ENHANCING INTERPRETABILITY OF PATIENT REPORTED OUTCOMES

ENHANCING METHODS FOR ANALYZING AND INTERPRETING

PATIENT-REPORTED OUTCOMES IN CLINICAL RESEARCH AND

EVIDENCE-BASED DECISION MAKING

By Tahira Devji, B.Sc

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfillment of the Requirements for

the Degree Doctor of Philosophy

McMaster University DOCTOR OF PHILOSOPHY (2019)

Hamilton, Ontario (Health Research Methodology)

TITLE: Enhancing methods for analyzing and interpreting patient-reported

outcomes in clinical research and evidence-based decision making

AUTHOR: Tahira Devji, B.Sc.

SUPERVISOR: Dr. Gordon H Guyatt, Distinguished Professor

NUMBER OF PAGES: xv, 206

ABSTRACT

In deciding whether to use a particular treatment for conditions such as depression, arthritis, or heart disease, clinicians and patients must balance the benefits against the side effects and burden. To make this trade-off, they must understand the likely degree of benefit in patients' symptoms and perceived wellbeing, best undertaken using patient-reported outcomes (PROs). PROs are measures of any aspect of a patients' health status that are obtained from direct patient inquiry without interpretation by a clinician or anyone else. PRO measures (PROMs) are increasingly used in clinical trials and systematic reviews to evaluate health care interventions, and information obtained from PROMs can guide clinical decisions and inform shared-decision making. The use of PROMs, however, involves challenges, the most important of which is deciding if a particular treatment effect is trivial, small but important, moderate or large. One way to make this judgment is to consider the minimal important difference (MID), the smallest change in a PROM score that is important enough that patients would consider a change in treatment to achieve that benefit. The number of published studies providing anchor-based MIDs for PROMs has grown rapidly over the last three decades, and researchers have proposed several anchor-based methods to derive MID estimates, each with its own merits and limitations. This thesis begins with the development of a framework to determine the extent to which the design and conduct of studies measuring anchor-based MIDs are likely to have protected against misleading estimates. Subsequently, this thesis presents a comprehensive inventory of empirically estimated anchor-based MIDs and their associated credibility for all PROMs published in the medical literature. Further, this thesis highlights critical issues that key stakeholders should consider, and demonstrates how the use of credible MIDs may inform the development of a clinical practice guideline in which PROs were identified as critically important. Finally, this thesis concludes with insights to improve the methodological quality and transparency for researchers in the PRO and MID field.

ACKNOWLEDGMENTS

First and foremost, I would like to thank Dr. Gordon Guyatt for his outstanding mentorship and for sharing his incredible passion towards teaching and life-long learning. Gordon, without your support and guidance and the opportunities you have afforded me, I could not have accomplished all that I have thus far in my academic career. You have nurtured me both professionally and personally, and I am forever indebted to you. I very much look forward to embarking on my post-doctoral journey with you and am thrilled to continue working on minimal important differences (a field you pioneered nearly 30 years ago – before I was born, remember?). Godspeed.

To my supervisory committee, Drs. Mohit Bhandari, Bradley Johnston, and Lehana Thabane, you have been invaluable sources of knowledge, expertise and inspiration. Thank you for the insightful feedback on this work and for providing me with advice that I will take with me throughout my career. Mo, you have uplifted and inspired my professional career tremendously. You put the wind in my sails to change the direction of my PhD research to focus on health status measurement. This encouragement, which came after a few years of supporting my Master's training, was truly a reflection of your great mentorship and commitment to helping your students find success and fulfillment in their professional careers. Today, you continue to create opportunities for me and provide guidance on navigating and achieving my career goals. I am fortunate to have you in my corner.

It has been an honour and privilege to complete my PhD training in the Health Research Methodology (HRM) program at the birthplace of evidence-based medicine among world-renowned experts. The Department of Health Research Methods, Evidence and Impact (HEI), which is known for its rich intellectual environment, impressive research expertise and breadth of disciplines, and an open inclusive approach to research and education, has exceeded all my expectations and

is truly an exceptional institution to complete graduate training. The Department of HEI also strongly encourages and facilitates interdisciplinary collaboration. Leveraging this culture during my PhD training, I have had countless opportunities to work with various groups on several projects with colleagues both locally and from around the world. The department and the program are products of a culture that is co-constructed by its greatest resource, its people. Thank you to all of the wonderful faculty, staff, and students in the Department of HEI and in the HRM program. I would also like to thank McMaster University, the Canadian Institutes of Health Research, and the Ontario Graduate Scholarship Program for providing me with salary and research support to make this work possible.

To all of the wonderful co-authors that I have had the privilege of collaborating with on all of my projects, I am grateful for your unwavering commitment, arduous efforts and insightful suggestions. In particular, Anila Qasim and Mark Phillips, your superb intellect and witty humour has made working with both of you an absolutely delightful experience. Your contributions have been instrumental to this work. Thank you.

To my dear colleague and friend, Dr. Alonso Carrasco-Labra, thank you for the wonderful partnership we have formed leading this important program of research together. I came to you nearly four years ago with an idea for my independent study as part of my comprehensive examination and then a few months later you helped me make the tough and outlandish decision to change the focus of my PhD research. This was best decision I have made for my career. Thank you for encouraging this decision and helping me realize my potential.

Once you are in the 'game' and demonstrate you are capable, driven and independent, it becomes easier to make connections, collaborate, and create opportunities within your field of work. Getting a foot 'in', however, is often the

most challenging step. I have two people to thank for bringing me into the realms of medicine and research: my cousin, Dr. Ally Prebtani, and Dr. Cindy Hamielec. Ally, you have mentored me from a young age and I am immensely grateful for your guidance in my personal and professional development. Thank you for all that you have done and continue to do for me; you are a second big brother to me. Cindy, you entrusted me with spearheading your critical care research program when I was only an undergraduate student, an experience that was crucial to my admittance and success in the HRM program. Thank you for helping me find my passion, which has had everything to do with where I've ended up today.

To my parents, Mumtaz and Iqbal, and my brother, Shafin, thank you for your constant love and encouragement throughout my 10 years of post-secondary education. Mom and Shafin, your willingness to sacrifice and relieve me of responsibilities so that I could pursue my educational ambitions without distraction or difficulty was the greatest gift of my life. Thank you for your selflessness and generosity. I am eternally grateful.

To my second parents, Rossmina and Nazmu, I am so lucky to have both in my life. You have truly kept me as your own daughter and shown me so much love. Thank you for giving me a second home in Mississauga, which made my long journeys from Markham to Hamilton that much more bearable.

My sweet and precious furry little friend, Phineus, you have been the best company during the good and challenging times. Thank you for your unconditional love and emotional support, for reminding me to take breaks (to pet you) by nudging my hand away from my laptop, for allowing me to tuck my cold feet under your warm coat of fur while I worked, and for the endless hugs.

Last but not least, thank you to my fiancé, Faiz, who has supported me through the

most important milestones in my life. Thank you for listening to me explain my research in far too much detail, for always letting me be the passenger during car rides so that I could work, take meetings, or sleep, for your understanding and patience with my demanding work schedule, and for genuinely taking interest in what I do. My love for you is immeasurable and I am incredibly fortunate to have you by my side.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

Abbreviation	Definition
ADL	Activities of daily living
APPADL	Ability to Perform Physical Activities of Daily Living Questionnaire
BMJ	British Medical Journal
CIHR	Canadian Institutes of Health Research
CI	Confidence Interval
CONSORT	Consolidated Standards of Reporting Trials
CINAHL	Cumulative Index of Nursing and Allied Health Literature database
ECOG	Eastern Cooperative Oncology Group
EQ-5D	EuroQol Five Dimensions Questionnaire
FACT-B	Functional Assessment of Cancer Therapy-Breast Cancer
FDA	Food and Drug Administration
GP	General practitioner
GRADE	Grading of Recommendations Assessment, Development and Evaluation
GROC	Global rating of change
HAQ	Health Assessment Questionnaire
HRQoL	Health-related quality of life
ICMJE	International Committee of Medical Journal Editors
KT	Knowledge Translation
KOOS	Knee injury and Osteoarthritis Outcome Score
MD	Mean difference
MID	Minimal important difference
NR	Not reported
NRS	Numerical rating scale
NSAID	Non-steroidal anti-inflammatory drugs
OA	Osteoarthritis
ODI	Oswestry Disability Index

OKS	Oxford Knee Score
PASS	Patient acceptable symptom state
PCOR	Patient-centered outcomes research
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRO	Patient reported outcome
PROQOLID	Patient Reported Outcome and Quality of Life Instruments Database
PROM	Patient reported outcome measure
PROMIS	Patient-Reported Outcomes Measurement Information System
RCT	Randomized clinical trial
ROC	Receiver operating characteristic
RTI	Raw transition item
SD	Standard deviation
SEM	Standard error of measurement
SF-36	36 item Short Form Survey
SMD	Standardized mean difference
TKR	Total knee replacement
VAS	Visual analogue scale
WOMAC	Western Ontario and McMaster University Osteoarthritis Index

DECLARATION OF ACADEMIC ACHIEVEMENT

Chapter 1: This chapter is unpublished. TD is the sole author.

Chapter 2: This chapter is in review at the *BMJ*. TD, ACL, GHG, BCJ, GN, SE conceived the study idea; TD, ACL, AQ, MP, GG led the development of the credibility instrument; TD, ACL, AQ, MP, ND, DZ, MB, XJ, RBP, OU, FF, SS, HPH, RWMV, HH, YR, RAS, and LL extracted data and assessed the credibility of MIDs in our inventory for the reliability analyses; TD and ACL wrote the first draft of the manuscript; TD, ACL, GG, AQ, MP, ND, DZ, RBP, OU, SS, HPH, RWMV, LL, BCJ, DLP, SE, TF, GN, HJS, MB, LT interpreted the data analysis and critically revised the manuscript.

Chapter 3: This chapter is in review at the *BMJ*. TD, ACL, BCJ, GN, SE, GHG conceived the study idea; TD, ACL, AQ, MP, GG created the data extraction form for the MID inventory and led the development of the credibility instrument; TD, ACL, AQ, MP, ND, DZ, MB, XJ, RBP, OU, FF, SS, HPH, RWMV, HH, YR, RAS, and LL extracted data and assessed the credibility of MIDs in our inventory; TD and ACL wrote the first draft of the manuscript; TD, ACL, GG, AQ, MP, ND, DZ, RBP, OU, SS, HPH, RWMV, LL, BCJ, DLP, SE, TF, GN, HJS, MB, LT interpreted the data analysis and critically revised the manuscript.

Chapter 4: This chapter is in review at the *BMJ*. TD wrote the first draft of the manuscript; TD, ACL, GG interpreted the data analysis and critically revised the manuscript.

Chapter 5: This chapter is published in *BMJ Open*. TD, GHG, ACL, RACS and POV conceived the study idea. TD and ACL performed the literature search. TD, ACL and GHG, among other colleagues, developed the credibility tool (core criteria) used in this study. TD performed the data analysis. TD, GHG, RWP, RB

and RACS interpreted the data analysis. TD wrote the first draft of the manuscript. TD, LL, BS and FF acquired the data and performed credibility assessments. TD, GHG, ACL, POV, RWP and RBP critically revised the manuscript.

Chapter 6: This chapter is unpublished. TD is the sole author.

Chapter 1: Introduction of the thesis

Evidence-based clinical and health policy decision-making has traditionally relied on measures of survival, longevity and major morbid events. Over the last 30 years, however, patient-centered outcomes research (PCOR) has emerged as a salient topic in evidence-based health care, in the process stressing the importance of patients' perspectives, interests and values. The central role of the patient perspective in PCOR mandates a greater emphasis on aspects of health-related of quality of life, including patients' symptoms, functional status, and perceived well-being. Questionnaires focusing on health status from the patients' own perspective - commonly referred to as patient-reported outcomes (PROs) – have gained prominence as an important strategy for determining the effect of interventions.

PROs are measures of any aspect of a patient's health status that are obtained from direct patient inquiry without interpretation by a physician or anyone else[1]. Although clinical and physiological measurements remain important markers of disease, illness, injury and their trajectories, PROs provide patients' perspectives on treatment benefits and harms, and are often the outcomes of most importance to patients. Indeed, information from PROs enhances patient-clinician communication, thereby promoting patient engagement in shared decision-making while improving patient satisfaction with care[2-6].

The PRO literature has grown exponentially over the last five decades (**Figure 1**) with several instruments (e.g. Short-Form 36, Beck Depression Inventory, Chronic Respiratory Questionnaire) in routine use in clinical research. The establishment of the National Institutes of Health's Patient-Reported Outcomes Measurement Information System (PROMIS)[7], guidance on the use of PRO measures (PROMs) from the U.S. Food and Drug Administration (FDA)[8], the Consolidated Standards of Reporting Trials (CONSORT) Extension for reporting of PROs in randomized trials[9], and integration of PROMs in primary and secondary care settings[10-15], all testify to the growing importance of PROMs in the assessment of clinical trial and

health system outcomes to inform patient-centered care, clinical decision-making
and health policy. Moreover, clinical trials and systematic reviews evaluating the
impact of interventions on PROMs and practice guidelines considering PROMs
have, over the last three decades, markedly increased[16].
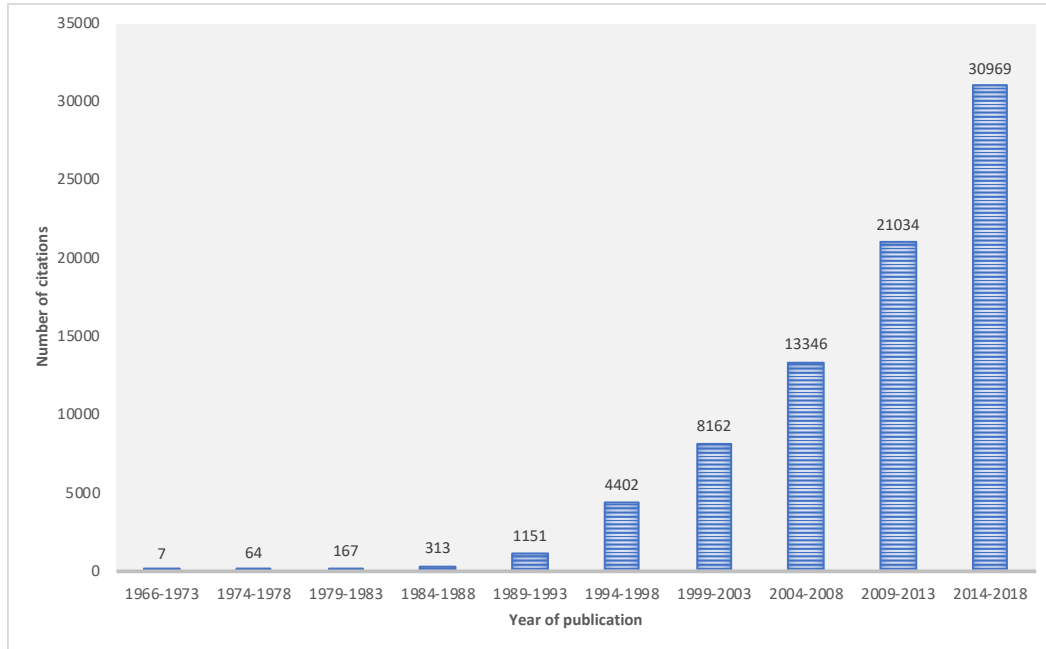
Figure 1. Number of publications found in PubMed with the search term 'patient
reported outcome', by 5-year stratum

With the growing body of clinical research relying on PROMs, decision-makers
need to judge the magnitude of treatment effects on these outcomes. The use of
PROMs, however, involves challenges, the most important of which is
interpretation of their results. Interpretability addresses how one might determine
differences in PROM scores that constitute trivial, small but important, moderate
or large differences in effect[17]. For instance, if a new pharmacological therapy to
treat major depression in adults improves patients' score on the Beck Depression
Inventory by three points relative to control, what are we to conclude about the
effectiveness of the new treatment? Is the treatment effect large or is it trivial?

Recognition of this challenge and potentially serious limitation of PROMs has led
to increased emphasis on the interpretation of differences based on patient-
importance[18,19]. The choice of what is considered as an important change in the
PROM score influences judgments about the design of clinical trials, the
interpretation of meta-analysis for the purpose of decision-making and guideline
recommendations, and the ultimate impact of a health care intervention[20].

There are various ways interpretation of PROMs can be facilitated, including the
consideration of the magnitude of effect in relation to the range of possible
instrument scores, and referring to the wide experience of clinicians using the
instrument in their clinical practice. These methods have severe limitations. The
most commonly used reference point for PROM interpretation is the minimal
important difference (MID), which is the smallest change in the outcome of interest,
either beneficial or harmful, that patients would perceive as important[20,21].

Knowledge of the MID facilitates the interpretation of treatment effects between
interventions compared in clinical trials, and systematic reviews and meta-analysis
of clinical trials, allowing decision-makers to discern whether or not individual
patients have obtained important benefits[22-26]. For instance, a guideline developer
using the Grading of Recommendations, Assessment and Evaluation (GRADE)
approach[27] to rate the quality of evidence and strength of a recommendation might
apply the MID as a threshold to decide whether the quality of evidence for a
particular PRO should be rated down for imprecision depending on whether the
confidence interval around a pooled effect estimate includes the MID, or if it in fact
lies well beyond the MID threshold[28] (**Figure 2**). The interpretation may be further
enhanced by conducting a responder analysis, which involves calculating the
proportion of patients in the intervention and control who have improved (or
deteriorated) by a score larger than the MID[29-31].
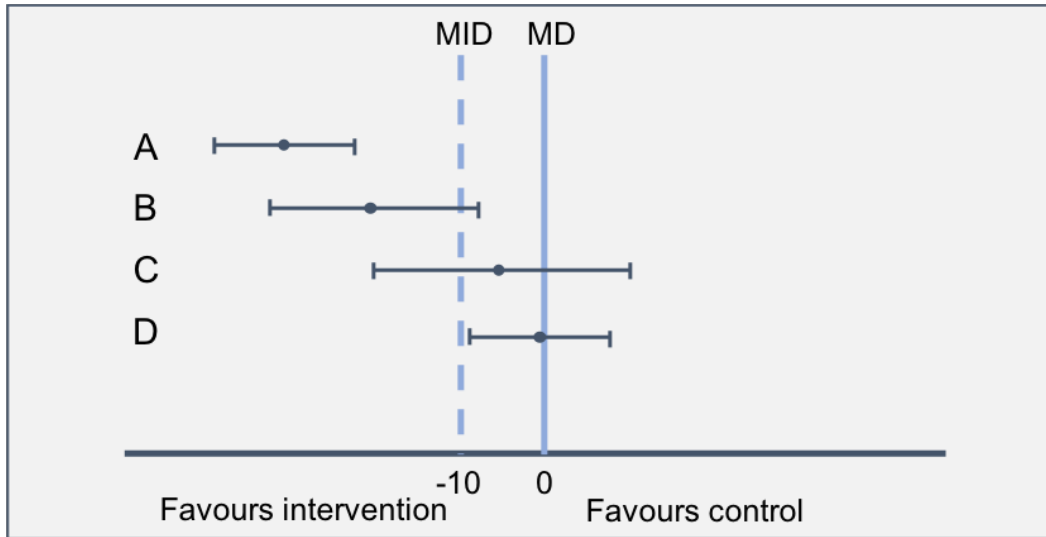
Figure 2. Hypothetical Example of change in PROM scores from four studies
presented in relation to an MID decision threshold

There are two main approaches for estimating an MID: distribution-based and
anchor-based methods. Distribution-based methods involve evaluating change in
the PROM in relation to either the probability that this change occurred by chance,
sample variation and measurement precision[29]. Common distribution-based
methods to estimate the MID include the use of the standard error of measurement
(SEM), a standard deviation (SD) of 0.5[30], the "effect size" (the effect in SD units,
with suggestions that treatment effects including 0.2 and 0.5 are minimally
important)[31]. Other distribution-based methods include the standardized response
mean[32], the responsiveness statistic[33], the paired *t-statistic*[34], growth curve
analysis[35], and the reliable change index[36] (**Appendix 1**). Distribution-based
approaches rely solely on the statistical characteristics of PROM scores and do not
reflect the patients' perspective, severely limiting these methods in aiding
interpretation of results.

In the anchor-based method, the PROM is compared to an independent standard –
the anchor – that is itself understandable and relevant to patients (e.g. measures of

symptoms, disease severity or response to treatment)[37]. The MID is calculated by associating the change in the PROM with an improvement or deterioration captured by the anchor. One common anchor-based approach to estimate the MID is the use of global transition questions (e.g. "Compared to 6 months ago before you had surgery, are you feeling better or worse – if so, to what extent?"). Other independent standards include preference ratings; comparisons to disease-related criteria, non-disease-related criteria, or known population(s); prognosis of future events; and changes in disease-related outcomes[38] (**Appendix 1**).

In the last three decades, the number of published studies providing anchor-based MIDs for PROMs has grown rapidly (**Figure 3**), and clinical trialists, systematic review authors, and guideline panellists are increasingly using MIDs to guide interpretation of treatment effects. Given the eruption of knowledge regarding MIDs and widespread recognition of their usefulness, the aim of this thesis is to address the use of MIDs for enhancing the interpretation of PROMs in clinical research, with a particular focus on the current state of the anchor-based MID literature and advancing methods for MID estimation.
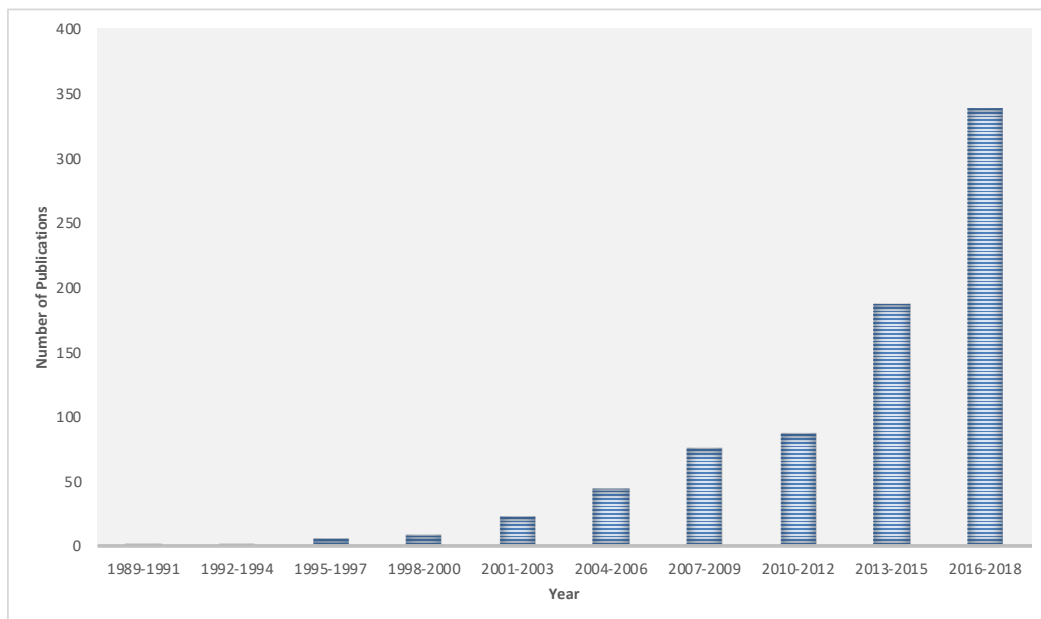
Figure 3. Number of published anchor-based MID estimation studies from 1989 up to October 2018 during each 3-year stratum

Chapter 2 documents the creation of a novel instrument – the first of its kind – to evaluate the credibility of empirically ascertained anchor-based MID estimates, and reports on the instrument's inter-rater reliability. Credibility refers to the extent to which the design and conduct of studies measuring MIDs are likely to have protected against misleading estimates.

Chapter 3 reports the development of a comprehensive inventory to systematically summarize all available published anchor-based MID estimates for PROMs in the medical literature. The chapter presents a complete exposition of the clinical context in which the MIDs were estimated, the specific anchor-based MID methodology applied, and an assessment of credibility for all MID estimates included in the inventory using the instrument developed in Chapter 2.

Chapter 4, informed by information from Chapter 2 and 3, elucidates the most critical issues of which MID investigators and users should be aware. These issues include the profusion of varying terms representing the MID concept; the multitude of diverse methods for MID estimation yielding different estimates and limited understanding about whether one or more represent better choices; and the urgent need for the development of a reporting guideline for anchor-based MID estimation studies to improve the completeness and transparency of MID reports and help achieve higher methodologic standards of MID estimation.

Chapter 5 provides a practical example demonstrating the value of the MID concept in facilitating interpretation of apparent treatment effects on critical outcomes of interest, including knee pain, function and health-related quality of life (HRQoL), in a clinical practice guideline addressing the impact of arthroscopic

surgery versus non-operative management in patients with degenerative knee disease. This systematic review identified credible MID estimates for the PROMs that informed evidence presentation in the associated systematic review of treatment effects, and judgments in the BMJ Rapid Recommendation. The guideline panel, aware through use of the MID that benefits associated with arthroscopy were very small, made a strong recommendation against knee arthroscopy.

This thesis ends with **Chapter 6,** which is a discussion of all of the previous chapters' findings, summarizing the most important findings, addressing limitations, and an exploration of future opportunities and directions for anchor-based MID determinations.

References

1. Food, Drug A. Guidance for industry: Patient-reported outcome measures: Use in medical product development to support labeling claims. Fed Regist. 2009;74(235):65132-65133.
2. Chen J, Ou L, Hollis SJ. A systematic review of the impact of routine collection of patient reported outcome measures on patients, providers and health organisations in an oncologic setting. BMC Health Serv Res. 2013;13:211.
3. Marshall S, Haywood K, Fitzpatrick R. Impact of patient-reported outcome measures on routine practice: a structured review. Journal of evaluation in clinical practice. 2006;12(5):559-568.
4. Santana MJ, Feeny D. Framework to assess the effects of using patient-reported outcome measures in chronic care management. Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation. 2014;23(5):1505-1513.
5. Valderas JM, Kotzeva A, Espallargues M, et al. The impact of measuring patient-reported outcomes in clinical practice: a systematic review of the literature. Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation. 2008;17(2):179-193.
6. Wasson JH, Stukel TA, Weiss JE, Hays RD, Jette AM, Nelson EC. A randomized trial of the use of patient self-assessment data to improve community practices. Effective clinical practice : ECP. 1999;2(1):1-10.
7. Cella D, Yount S, Rothrock N, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS): Progress of an NIH Roadmap Cooperative Group During its First Two Years. Med Care. 2007;45(5 Suppl 1):S3-s11.
8. US Food and Drug Administration (FDA). Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. Health and Quality of Life Outcomes. 2006;4:1-20.
9. Calvert M, Blazeby J, Altman DG, et al. Reporting of patient-reported outcomes in randomized trials: The consort pro extension. JAMA. 2013;309(8):814-822.
10. Crandall W, Kappelman MD, Colletti RB, et al. ImproveCareNow: the development of a pediatric inflammatory bowel disease improvement network. Inflammatory bowel diseases. 2010;17(1):450-457.
11. Devlin NJ, Appleby J. Getting the most out of PROMs. Putting health outcomes at the heart of NHS decision-making London: The King's Fund. 2010.
12. Devlin NJ, Parkin D, Browne J. Patient-reported outcome measures in the NHS: new methods for analysing and reporting EQ-5D data. Health economics. 2010;19(8):886-905.
13. Haverman L, Engelen V, van Rossum MA, Heymans HS, Grootenhuis MA. Monitoring health-related quality of life in paediatric practice: development of an innovative web-based application. BMC pediatrics. 2011;11(1):3.

14. Hvitfeldt H, Carli C, Nelson EC, Mortenson DM, Ruppert BA, Lindblad S. Feed forward systems for patient participation and provider support: adoption results from the original US context to Sweden and beyond. Quality Management in Healthcare. 2009;18(4):247-256.

15. Wasson JH, Benjamin R, Johnson D, Moore LG, Mackenzie T. Patients use the Internet to enter the medical home. The Journal of ambulatory care management. 2011;34(1):38-46.

16. Johnston BC, Ebrahim S, Carrasco-Labra A, et al. Minimally important difference estimates and methods: a protocol. BMJ Open. 2015;5(10):e007953.

17. Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. Annals of internal medicine. 1993;118(8):622-629.

18. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Controlled clinical trials. 1989;10(4):407-415.

19. Johnston BC, Thorlund K, Schunemann HJ, et al. Improving the interpretation of quality of life evidence in meta-analyses: the application of minimal important difference units. Health Qual Life Outcomes. 2010;8:116.

20. Schünemann HJ, Guyatt GH. Commentary—Goodbye M(C)ID! Hello MID, Where Do You Come From? Health Serv Res. 2005;40(2):593-597.

21. Schunemann HJ, Puhan M, Goldstein R, Jaeschke R, Guyatt GH. Measurement properties and interpretability of the Chronic respiratory disease questionnaire (CRQ). Copd. 2005;2(1):81-89.

22. Burris HA, 3rd, Lebrun F, Rugo HS, et al. Health-related quality of life of patients with advanced breast cancer treated with everolimus plus exemestane versus placebo plus exemestane in the phase 3, randomized, controlled, BOLERO-2 trial. Cancer. 2013;119(10):1908-1915.

23. Guyatt G, Schunemann H. How can quality of life researchers make their work more useful to health workers and their patients? Qual Life Res. 2007;16(7):1097-1105.

24. Katz NP, Paillard FC, Ekman E. Determining the clinical importance of treatment benefits for interventions for painful orthopedic conditions. J Orthop Surg Res. 2015;10.

25. Nonoyama ML, Brooks D, Guyatt GH, Goldstein RS. Effect of oxygen on health quality of life in patients with chronic obstructive pulmonary disease with transient exertional hypoxemia. American journal of respiratory and critical care medicine. 2007;176(4):343-349.

26. Weinblatt ME, Kavanaugh A, Genovese MC, et al. Effects of fostamatinib (R788), an oral spleen tyrosine kinase inhibitor, on health-related quality of life in patients with active rheumatoid arthritis: analyses of patient-reported outcomes from a randomized, double-blind, placebo-controlled trial. J Rheumatol. 2013;40(4):369-378.

27. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. Bmj. 2008;336(7650):924-926.

28. Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines 6. Rating the quality of evidence--imprecision. J Clin Epidemiol. 2011;64(12):1283-1293.

29. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. J Clin Epidemiol. 2003;56(5):395-407.

30. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. Medical care. 2003;41(5):582-592.

31. Cohen J. Statistical power analysis for the behavioral sciences. Academic press; 2013.

32. Stucki G, Liang MH, Fossel AH, Katz JN. Relative responsiveness of condition-specific and generic health status measures in degenerative lumbar spinal stenosis. J Clin Epidemiol. 1995;48(11):1369-1378.

33. Guyatt GH, Bombardier C, Tugwell PX. Measuring disease-specific quality of life in clinical trials. CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne. 1986;134(8):889-895.

34. Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness: a critical review and recommendations. Journal of clinical epidemiology. 2000;53(5):459-468.

35. Speer DC, Greenbaum PE. Five methods for computing significant individual client change and improvement rates: support for an individual growth curve approach. Journal of consulting and clinical psychology. 1995;63(6):1044.

36. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. Journal of consulting and clinical psychology. 1991;59(1):12.

37. Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR. Methods to explain the clinical significance of health status measures. Mayo Clinic proceedings. 2002;77(4):371-383.

38. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. Journal of Clinical Epidemiology. 2003;56(5):395-407.

Appendix 1. Empirical methods to estimate the MID

Anchor-based methods

Type	Method	PROM evaluated in relation to:
Cross-sectional	Comparison to disease-related criteria	Disease severity or diagnosis
	Comparison to non-disease related criteria	Impact of life events
	Preference ratings	Pairwise comparison of health states
	Comparison to known population(s)	Functional or dysfunctional populations
Longitudinal	Global ratings of change	Patients' or clinicians' ratings of improvement
	Prognosis of future events	Those experiencing and not experience some future event
	Changes in disease related outcome	Changes in clinical outcome

Table reproduced from Crosby RD. J Clin Epidemiol. 2003;56(5):395-407. PROM, patient reported outcome measure

Distribution-based methods

Method	Reference	PROM evaluated in relation to:	Calculation
Paired *t*-statistic	Husted et al., 2000	Standard error of the mean change	$\dfrac{x_1 - x_0}{\sqrt{\dfrac{\Sigma(d_1 - d)^2}{n(n-1)}}}$
Growth curve analysis	Speer and Greenbaum, 1995	Standard error of the slope	$\dfrac{B}{\sqrt{V}}$
Effect size	Cohen, 1988	Pre-test standard deviation	$\dfrac{x_1 - x_0}{\sqrt{\dfrac{\Sigma(d_1 - d)^2}{(n-1)}}}$
Standardized response mean	Stucki et al., 1995	Standard deviation of change	$\dfrac{x_1 - x_0}{\sqrt{\dfrac{\Sigma(d_1 - d)^2}{(n-1)}}}$
Responsiveness statistic	Guyatt et al., 1986	Standard deviation of change in a stable group	$\dfrac{x_1 - x_0}{\sqrt{\dfrac{\Sigma(d_{1\,stable} - d_{stable})^2}{(n-1)}}}$
Standard error of measurement	Wyrich et al., 1999	Standard error of measurement	$\dfrac{x_1 - x_0}{\sqrt{\dfrac{\Sigma(x_1 - x_0)^2}{(n-1)}}(\sqrt{1-r})}$
Reliable change index	Jacobson and Truax, 1991	Standard error of the measurement difference	$\dfrac{x_1 - x_0}{\sqrt{2(\text{SEM})^2}}$

Table reproduced from Crosby RD. J Clin Epidemiol. 2003;56(5):395-407. PROM, patient reported outcome measure

Chapter 2: Development and inter-rater reliability of an instrument to evaluate the credibility of anchor-based minimal important difference estimates for patient reported outcomes

Tahira Devji*, Alonso Carrasco-Labra*, Anila Qasim, Mark Phillips, Niveditha Devasenapathy, Dena Zeraatkar, Meha Bhatt, Xuejing Jin, Romina Brignardello-Petersen, Olivia Urquhart, Farid Foroutan, Stefan Schandelmaier, Hector Pardo-Hernandez, Robin WM Vernooij, Hsiaomin Huang, Yamna Rizwan, Reed Siemieniuk, Lyubov Lytvyn, Bradley C Johnston, Donald L Patrick, Shanil Ebrahim, Toshi Furukawa, Gihad Nesrallah, Holger J Schunemann, Mohit Bhandari, Lehana Thabane, Gordon H Guyatt

*Co-first authorship

Submitted to: BMJ [Dec 2018]

ABSTRACT

Objective: Anchor-based approaches for minimal important difference (MID) estimation relate a change in a patient reported outcome measure (PROM) to an external criterion (i.e. the anchor) that is understandable and relevant to patients. The aim of this study was to develop an instrument to evaluate the credibility of anchor-based MID estimates for PROMs and assess the reliability of this instrument. We defined credibility as the extent to which the design and conduct of studies measuring MIDs are likely to have protected against misleading estimates.

Design: On the basis of a literature review and our groups' experience with methods of ascertaining MIDs, we developed initial criteria for evaluating the credibility of anchor-based MIDs. Iterative discussion among the team and pilot testing with experts in the field and potential users led to the development of the final version of the instrument. Teams of two reviewers independently applied the newly developed instrument to evaluate credibility of a random sample of MID estimates for inter-rater reliability testing of the instrument.

Main outcomes and measures: Core credibility criteria applicable to all anchor types, additional criteria for transition rating anchors, and inter-rater reliability coefficients.

Results: The credibility instrument includes the following core criteria relevant for any anchor: the anchor is rated by, interpretable, and relevant to the patient; the MID estimate is precise; the correlation between the anchor and PROM is satisfactory, and the authors select a threshold on the anchor that reflects a small but important difference. The extension for transition rating anchors includes the following items: the time elapsed between baseline and follow-up measurement for MID estimation is optimal; and the correlations of the transition rating with the pre, post, and change score in the PROM are satisfactory. The inter-rater reliability for

all of the core criteria and the single evaluable criterion from the extension ranged
from good (Cohen's kappa ≥0.7) to very good (≥0.8) agreement. Reporting issues
prevented us from evaluating reliability of the three remaining criteria in the
extension for transition rating anchors.

Conclusions: Researchers, clinicians, trialists and health care policy decision-
makers can now make use of a reliable instrument to evaluate the design, conduct
and analysis of studies estimating anchor-based MIDs.

INTRODUCTION

For decades, evaluation of outcomes in clinical research and practice has relied on survival, longevity, major morbid events (e.g. mortality, stroke) and laboratory endpoints (e.g. serum creatinine, hemoglobin A1C). More recently, a shift towards patient-centered care has resulted in a greater emphasis on evaluating patients' symptoms, functional status, and perceived well-being. These outcomes, typically measured from direct patient inquiry using questionnaires – previously referred to as 'health-related quality of life' measures – are now most commonly labelled as patient-reported outcomes (PROs). PROs represent reports of patients' health status that comes directly from patients without interpretation by a physician or anyone else[1].

Many PRO measures (PROMs) have established validity, reliability and responsiveness. The interpretation of PROMs has, however, remained challenging. In particular, clinical application requires knowing if an apparent treatment effect is trivial in magnitude, small but important, moderate or large[2]. To aid interpretation of PROMs, researchers developed a concept known as the minimal important difference (MID)[3]. The MID, which provides a measure of the smallest change – either positive or negative – that patients perceive as an important benefit or harm[3,4], represents the most commonly used reference point for PROM interpretation.

There are two approaches for determining the MID: distribution- and anchor-based methods. Distribution-based methods rely on the statistical characteristics of the distribution of PROM scores and thus fail to incorporate patients' perspective, severely limiting their usefulness in aiding interpretation of PROMs[5,6]. Anchor-based methods address the MID by associating a PROM with an independent measure – an external criterion or "anchor" – that is understandable and relevant to patients[7], and are accepted as the optimal way of establishing the MID.

Anchor-based MID estimations vary in the choice of anchor, the relation between the anchor and PROM under consideration, the statistical methods used to establish the MID, and study sample size. Some of these choices are more satisfactory than others – indeed, poor choices can lead to MIDs that mislead, and misleading MIDs will result in seriously flawed interpretation of results. Thus, for optimal use of MIDs, investigators and decision makers must be able to distinguish between more and less credible MIDs.

We define credibility as the extent to which the design and conduct of studies measuring MIDs are likely to have protected against misleading estimates. Currently, no accepted standards for appraising the credibility of an anchor-based MID exist. In this article, we describe the development of an instrument to evaluate the credibility of anchor-based MIDs and report on the inter-rater reliability of this instrument.

METHODS

Development of a Credibility Instrument for Studies Determining MIDs

Item generation

In a related article, we reported on the methods and results of a systematic survey to develop an inventory of all published anchor-based MIDs for PROMs in the medical literature (Submitted Dec 2018 to the BMJ). Briefly, we searched MEDLINE, EMBASE and PsycINFO for studies published from 1989 to April, 2015[3]. The search strategy, adapted to each database, included terms representing the MID concept along with terms addressing PROMs (**Appendix 2**). From the search results, we identified and reviewed methods articles addressing MID estimation using anchor-based approaches, including theoretical descriptions, summaries, commentaries and critiques. We used standard thematic analysis techniques[8] to abstract concepts related to the credibility of studies estimating

MIDs, specifically the extent to which the design, conduct and analysis of studies are likely to have protected against misleading estimates.

On the basis of this survey of the literature and our groups' experience with methods of ascertaining MIDs[3,9-14], we developed initial criteria for evaluating the credibility of anchor-based MIDs.

Face and content validity

We presented the initial criteria to experts (i.e. researchers with expertise in instrument development, MID estimation and PROs) and target users (i.e. clinicians, trialists, systematic reviewers and guideline developers). These individuals reviewed the instrument for clarity, wording, comprehensiveness and item relevance, and provided suggestions to improve the instrument; we incorporated this feedback. An early version of the instrument has been published elsewhere[15]. Subsequent work, including application of the draft instrument to anchor-based MID estimation studies included in our MID inventory (Submitted Dec 2018 to the BMJ) and additional applications of the instrument to inform the development of a clinical practice guideline[16], led to item modification and reduction. We conducted this iterative process of pilot testing and user feedback until we achieved consensus for the final version of the credibility instrument.

Response options

With the exception of the first item, which has a yes/no response, each item provides a five-point adjectival scale. The response options for items in the instrument are: definitely yes; to a great extent; not so much; definitely no; impossible to tell, with wording such that a response of 'definitely yes' indicates no concern regarding the credibility of the MID estimate. Responses of 'definitely yes' and 'definitely no' imply that information provided in the MID report under evaluation allows an unequivocal judgment in relation to the item; the "to a great

extent" and "not so much" responses denote lower certainty. In the absence of information or sufficient detail to make an informed judgment about credibility, one may use the "impossible to tell" response option.

Reliability Study of the Credibility Instrument

Sample of MID estimates and Raters

In our aforementioned inventory of anchor-based MIDs, we summarized over 3,000 estimates and their associated credibility, including MIDs for PROMs across different populations, conditions, and interventions, obtained using different anchors and statistical methods (Submitted Dec 2018 to the BMJ). We enlisted help from Masters and PhD trainees with background in health research methodology to conduct study screening, data extraction and the credibility assessment. Prior to commencing the review process, the reviewers received extensive training regarding MID methodology, including background readings of key MID methods articles, web teleconferences to review screening and data extraction materials, and pilot and calibration exercises. Teams of two reviewers independently extracted relevant data from included studies for each MID estimate, collecting information on study design, characteristics of the PROM, anchor and analytic method, sample size, the MID estimate and associated measure of precision, time elapsed between administration of the PROM and follow up assessments of the PROM and anchor (for longitudinal designs); and applied the newly developed instrument to evaluate credibility of the MID estimates.

Sampling method

For a random sample of 200 MID estimates from our inventory, we retrieved the credibility assessments performed by each pair of reviewers using the newly developed instrument. We sampled in excess (see sample size below) to account for potential discrepancies in the MIDs extracted between reviewers and incomplete data. For instance, situations in which one reviewer could have missed an MID

reported in the study, we would only have a single credibility assessment. To ensure
observations in our sample were independent of each other, when a single study
reported multiple MIDs, we only included one estimate.

Sample size

We tested the reliability of our credibility instrument using classical test theory[17].
Given that assessments regarding credibility involve subjective judgments and
different individuals collecting data may experience and interpret phenomena of
interest differently, we measured inter-rater reliability. According to Shoukri[18],
considering 2 raters per MID estimate, an expected reliability of 0.7, with a desired
95% confidence interval (CI) width of 0.2, and an α of 0.05, would require a
minimum of 101 MIDs assessed per rater.

Analysis

For each item of the instrument, we calculated inter-rater reliability and associated
95% CI, as measured by a weighted kappa, κ, with quadratic weights assigned using
the formula: $w_i = 1 - \frac{i^2}{(k-1)^2}$, where i is the difference between categories (i.e.
response options) and k is the total number of categories. We considered a reliability
coefficient of at least 0.7 to represent good inter-rater reliability[19-21].

RESULTS

We identified 41 relevant MID methods articles[4,5,7,12,13,22-57] that informed the item
generation stage of instrument development. There were two substantive
modifications from the first draft[15] to the definitive instrument presented here. In
the first, we removed three items due to issues of redundancy and relevance; re-
phrased one item addressing to what extent the anchor and the PROM are measuring
the same construct; and added one new item addressing the precision around the
MID estimate. In the second, we added a new item evaluating whether the anchor
threshold selected for MID estimation reflects a small but important difference; and

developed additional criteria for assessing the credibility of a transition rating anchor (further described below).

Credibility Instrument

The instrument consists of five criteria essential for determining the credibility of any anchor-based MID (**Table 1**). In our inventory of anchor-based MIDs (Submitted Dec 2018 to the BMJ) and a separate systematic review to identify MIDs for knee specific PROMs[16], we found that MIDs are most often derived using transition rating anchors. Anchors of this sort require patients to recall a prior health state and compare that state to how they are currently feeling. This retrospection required criteria ensuring that transition ratings accurately reflect the change in health status and are not unduly influenced by the baseline or endpoint status; thus, for this context, we developed a four-item extension of the core credibility instrument (**Table 2**). Below, we describe each question included in the instrument followed by an explanation detailing the relevance of the item for evaluating credibility. We provide two worked examples in **Appendix 3** in which we have applied our instrument to assess the credibility of two MID estimates, each from a published study.

Table 1. Credibility instrument for judging the trustworthiness of minimal important difference estimates

MINIMAL IMPORTANT DIFFERENCE CREDIBILITY ASSESSMENT TOOL

Is the patient or necessary proxy responding directly to both the patient-reported outcome measure and the anchor? ☐ Yes ☐ No ☐ Impossible to tell	*If a clinician or anyone else is responding to the anchor directly and the patients are capable of providing this information, the answer should be "no." Any other necessary proxy (e.g. caregiver, parent, wife, relative) responding to the anchor, the answer is "yes".* **Rationale:**
Is the anchor easily understandable and relevant for patients or necessary proxy? ☐ Definitely yes ☐ To a great extent ☐ Not so much ☐ Definitely no ☐ Impossible to tell	*With "easily understandable and relevant" we mean that, when presented with the anchor (either actually presented or hypothetically) as an outcome, and without too much education, the patients are able to understand the data provided for the outcome (anchor) and use it easily for decision-making. For example, when addressing a multi-item patient-reported outcome measure addressing the potential therapeutic effects of an intervention for iron-deficiency anemia, an anchor of patient's global rating of improvement in fatigue may be easier to understand and more relevant for decision-making than serum iron levels.* *If you were a patient, how would you answer this question?* **Rationale:**
Has the anchor shown good correlation with the patient-reported outcome measure? ☐ Definitely yes (≥0.7) ☐ To a great extent (≥0.5 to <0.7) ☐ Not so much (≥0.3 to <0.5) ☐ Definitely no (<0.3) ☐ Impossible to tell	*This assessment is made using the correlation coefficients reported by the authors. If the anchor is a transition question then this is correlation between the transition item and the PROM change score. For any other anchor, this is the correlation between the change in the anchor and the change in the PROM. If the study is cross-sectional, this is the correlation between the anchor and the PROM score. Only consider the absolute value of the correlation coefficient.* **Rationale:**

Is the MID precise?

☐ Definitely yes (≤20% **or** ≥200 patients)

☐ To a great extent (21-50% **or** 150-199 patients)

☐ Not so much (51-100% **or** 100-149 patients)

☐ Definitely no (>100% **or** <100 patients)

☐ Impossible to tell

*Precision around the MID estimate is quantified by the width of the 95% CI and expressed as a percentage. For example, if the MID estimate is 23.5 and the 95% CI ranges from 23.1 to 23.8, then precision may be calculated as: 23.8 – 23.1 / 23.5 * 100 = 3%. According to our guide provided for our responses to this credibility question, a result of 3% would warrant a rating of definitely yes. In many cases, the authors may not report any measure of variability (e.g. SD, SE, 95% CI). In these situations, we ask that you consider the sample size used to estimate the MID. We provide ranges for both situations (i.e. percentage of the confidence interval width in relation to the MID, and sample sizes) to help inform your judgment. If the judgments according to the two criteria differ, we suggest using the higher (more permissive) of the two ratings.*

Rationale:

Does the threshold or difference between groups on the anchor used to estimate the MID reflect a small but important difference?

☐ Definitely yes

☐ To a great extent

☐ Not so much

☐ Definitely no

☐ Impossible to tell

Establishing the degree of change on a PROM that constitutes the MID requires some knowledge about the degree of change on the anchor that is small but important to patients. In addition to inspecting the threshold on the anchor, it is necessary to judge whether the method of analysis indeed calculates a small but important difference. Below, we present examples and provide associated guidance.

For transition rating anchors, consider the wording and number of responses. For instance, the mean change in PROM score in patients with a transition rating anchor scale designation of 'a little better' on a seven-point scale including the categories 'much worse, somewhat worse, a little worse, no change, a little better, somewhat better, much better,' as reflecting an MID would warrant a definitely yes, whereas a choice of "much better" would warrant a definitely no.

In some cases, authors may use a threshold for their analysis and include only patients who achieved this threshold; other times, they may include patients who achieved this threshold or greater. For instance, the investigators may define the MID as the mean change in the PROM score in patients who achieved a ≥5% change in weight loss. This approach includes even those patients who had a 10%, 30% or 50% reduction in weight loss and thus would warrant a definitely no.

Rationale:

PROM patient reported outcome measure; MID minimal important difference; CI confidence interval, SD standard deviation, SE standard error

Table 2. Credibility instrument extension for transition rating anchors

MINIMAL IMPORTANT DIFFERENCE CREDIBILITY ASSESSMENT TOOL – EXTENSION FOR TRANSITION RATINGS

| **Is the amount of elapsed time between baseline and follow-up measurement for MID estimation optimal?**
☐ Definitely yes (≤ 1 month)
☐ To a great extent (>1 to ≤2 months)
☐ Not so much (>2 months to ≤3 months)
☐ Definitely no (>3 months)
☐ Not reported | *If there is a range of follow-up reported, consider the following when making your judgment: If the range falls over 3 categories (e.g. 3 weeks to 3 months), then select the middle category (i.e. in this example, you would select 'to a great extent'); If the range falls over 2 categories (e.g. 6 weeks to 3 months), then select the more conservative option (longest follow-up) (i.e. in this example, you would select 'not so much')*
Rationale: |

To answer the next 3 questions, you first need to determine if the scale of the anchor and PROM are in the same direction.

*For each question we provide 2 guides: If higher values on the anchor and PROM represent the SAME state (i.e. both represent a better or worse condition), use **Guide A**; If higher values on the anchor and PROM represent DIFFERENT states (i.e. higher scores on the PROM are worse, while higher values on the anchor are better), use **Guide B***

| **Does the transition item have a substantial correlation with the PROM score at follow-up?**
☐ Definitely yes
☐ To a great extent
☐ Not so much
☐ Definitely no
☐ Not reported | **Guide A**
Definitely yes (>0.2)
To a great extent (0.1 to 0.2)
Not so much (<0.1)
Definitely no (negative correlation)
Rationale: | **Guide B**
Definitely yes (<-0.2)
To a great extent (-0.1 to -0.2)
Not so much (>-0.1)
Definitely no (positive correlation) |
| **Does the transition item correlate with the PROM score at baseline?**
☐ Definitely yes
☐ To a great extent
☐ Not so much
☐ Definitely no
☐ Not reported | *Definitely yes (negative correlation)*
To a great extent (<0.1)
Not so much (0.1 to 0.2)
Definitely no (>0.2)
Rationale: | *Definitely yes (positive correlation)*
To a great extent (>-0.1)
Not so much (-0.1 to -0.2)
Definitely no (<-0.2) |

Is the correlation of the transition item with the PROM change score appreciably greater than the correlation of the transition item with the PRO score at follow-up?	*Definitely yes (≥0.2)*	*Definitely yes (≤-0.2)*
	To a great extent (0.1 to < 0.2)	*To a great extent (-0.1 to >-0.2)*
☐ Definitely yes	*Not so much (0 to <0.1)*	*Not so much (0 to >-0.1)*
☐ To a great extent	*Definitely no (<0)*	*Definitely no (>0)*
☐ Not so much	**Rationale:**	
☐ Definitely no		
☐ Not reported		

PROM patient reported outcome measure; MID minimal important difference; CI confidence interval, SD standard deviation, SE standard error

Core criteria

Item 1. Is the patient or necessary proxy responding directly to both the patient reported outcome measure and the anchor?

An anchor-based method for estimating an MID involves linking a specific PROM (e.g. Short-Form 36, Beck Depression Inventory, Chronic Respiratory Questionnaire) to an external criterion such as a patient or physician transition rating, another PROM, or a clinical endpoint (e.g. hemoglobin level, Eastern Cooperative Oncology Group (ECOG) performance status). Patient-reported anchors are more desirable than clinical measures or those that are clinician assessed. Situations in which the patient is unable to directly provide information to inform the outcome (e.g. elderly individuals with dementia, infants and pre-verbal toddlers) require a proxy respondent. We suggest using the same standards recommended for a patient directly responding to the PROM when evaluating the credibility of MIDs for a necessary proxy-reported PROM.

Item 2. Is the anchor easily understandable and relevant for patients or necessary proxy?

A desirable anchor is one that is easily understandable and is highly relevant to patients. Typical appropriate anchors include global ratings of change on health status[11,58-60], status on an important and easily understood measure of function[61], the presence of symptoms[62], disease severity[63], response to treatment[63,64], or the prognosis for future events such as mortality[62,65,66], health care utilization[67] or job loss[62,68,69].

Item 3. Has the anchor shown a satisfactory correlation with the patient-reported outcome measure?

The usefulness of anchor-based approaches is critically dependent on the relationship between the PROM and the anchor. When determining the credibility of the MID, we consider how closely the anchor is related to the target PROM and

attribute greater importance to MIDs generated from closely linked concepts. That is, the anchor and PROM should be measuring the same or similar underlying constructs, and therefore should be appreciably correlated. A moderate to high correlation (at least 0.5) suggests the validity of the anchor[34,70,71]. An anchor that has very low or no correlation with the PRO instrument will likely yield inaccurate MID estimates. The instrument provides a guide for judging the correlation coefficient.

Item 4. Is the MID precise?

To judge precision, we focus on the 95% CI around the point estimate of the MID. When authors do not provide a measure of precision, the number of patients informing the MID estimation provides an alternative criterion. In the instrument, we provide a guide for judging precision when the investigators report the 95% CI around the MID estimate based on the likelihood that inferences regarding the magnitude of a treatment effect would differ at the extremes of the confidence interval. If a measure of precision is not reported, we provide guidance regarding appropriate sample size based on the relation between sample size and precision in studies in the inventory that did provide 95% CIs.

Item 5. Does the threshold or difference between groups on the anchor reflect a small but important difference?

To respond to this credibility question, one needs to make a judgement regarding whether the selected threshold, or groups compared on the anchor, reflect a small (rather than moderate or large) but important difference. Even after the threshold is set, there are a multitude of analytic methods to compute the MID, and it is necessary to judge whether the chosen method of analysis calculates an MID. **Box 1** provides a framework for making these judgments, and we provide some examples of high and low credibility MIDs estimated with different types of anchors.

Box 1. Judging whether the MID represents a small but important difference

> 1. What is the original scale of the anchor and is it transformed in any way?
> 2. Does the scale (or transformed scale) of the anchor capture variability in the underlying construct?
> 3. What is the threshold used or comparison being made on the anchor? Does this threshold/comparison represent a difference that is minimally important?
> 4. Does the analytical method ensure that the MID represents a small but important difference? Example 4 below demonstrates how a poorly chosen analytic method could yield misguided MID estimates.

Examples of high credibility:

1. Investigators calculated the MID for the Western Ontario and McMaster University Osteoarthritis Index (WOMAC) pain domain as the mean change in the WOMAC pain score in patients who reported themselves as "a little better" to the question "how was the pain in your operated hip during the past week, as compared to before the operation" offering response options extremely better, very much better, much better, better, a little better, a very little better, almost the same/hardly any better, no change (with parallel responses for worsening)[51].

2. To estimate the MID for the Functional Assessment of Cancer Therapy-Breast Cancer (FACT-B), investigators compared ECOG performance status (scores range from 0-4, higher scores signify worse performance status) at follow up to baseline performance status. If the rating at follow up was lower than at baseline, then patients were considered "improved"; if higher score, they were considered "worsened". A patient was considered "unchanged" if the scores at baseline and follow-up were the same. The MID was defined as the mean change in FACT-B scores among patients who were "improved"[72].

Examples of low credibility:

3. Patients responded to the following: "Compared to before treatment my back problem is a) much better, b) better, c) unchanged, d) worse". Investigators defined the MID for deterioration for the Oswestry Disability Index (ODI) by calculating the difference in ODI score between those patients who rated themselves "worse" and those who were "unchanged"[73]. This is low credibility because worse could mean a little worse or much worse.

4. Investigators estimated the MID for the Ability to Perform Physical Activities of Daily Living Questionnaire (APPADL) by taking the difference in mean APPADL change scores for those who achieve 5% or more weight loss from baseline to 6 months and those who achieved less than 5%[74]. This is problematic because we have no idea how patients whose weight falls by 6% react – that is, are they pleased they have made a substantial weight reduction, consider this small but important, or regard it as trivial. Further, the researchers use a misguided analytic method. In their group of patients who they classify as having a small but important improvement, they included not only patients who had a 5%, but also a 10%, 30% or 50% reduction in weight loss together. Subtracting the APPADL mean change score for the group of patients achieving a less than 5% change in weight loss from those that experienced a change greater than 5%, could yield an estimate for the MID that constitutes a small, moderate or even large difference depending on the proportion of patients who achieved large percentage weight losses.

Extension for Transition Rating Anchors
Item 1. Is the amount of elapsed time between baseline and follow-up measurement for MID estimation optimal?

Despite the intuitive appeal of transition questions, patients have considerable difficulty recalling prior health states[34,37,75]. As the duration of time over which patients must cast their memory increases, the difficulty increases[34,37]. Patients can often recall prior states for periods of up to 4 weeks[34]; as time intervals extend into months, patients are more likely to confuse change over time with current status[37].

Judgments for items 2-4 of the extension requires knowledge regarding the directionality of the PROM and transition scale. In the instrument, we provide guidance to address situations in which higher scores on both the PROM and anchor represent the same direction (i.e. both represent a worse or better condition) and when they represent different directions.

Item 2. Does the transition item have a substantial correlation with the PROM score at follow-up?

Ideally, the correlation between the transition rating with the pre-score and the transition rating with the post-score would be equal and opposite, an ideal that seldom occurs. To the extent that the post-score shows at least some correlation with the transition, the MID estimate is more credible than if there were no correlation[34].

Item 3. Does the transition item correlate with the PROM score at baseline?

If the pre-score correlates with the transition rating, we are more confident that patients are taking their baseline status into account when scoring the transition rating[34].

Item 4. Is the correlation of the transition item with the PROM change score appreciably greater than the correlation of the transition item with the PROM score at follow-up?

A correlation of at least 0.5 between the transition rating and the change in PROM is necessary but insufficient to confirm that the transition rating is in fact measuring change as opposed to current health status. A correlation of the post-score with the transition that is similar or greater than the correlation of the change with the transition provides evidence that the rating likely reflects only current status, and thus decreases confidence in the MID estimate[34].

The instrument provides a guide for judging the correlation coefficients addressed in items 2-4.

Reliability analyses

The analysis for the assessment of inter-rater reliability included 135 MIDs assessed by two raters for the core credibility criteria and 137 MIDs for the first item in the extension. Participants providing credibility ratings included Masters and PhD trainees with backgrounds in health research and MID methods. For the remaining items in the extension, only 12 studies reported the correlation between the post-score and transition rating addressed in item two and four, and 10 studies provided the correlation between the pre-score and transition rating required for item three. Due to the limited sample sizes we were unable to conduct an evaluation of the inter-rater reliability for these items.

Overall, the inter-rater reliability for all items ranged from good (Cohen's kappa ≥0.7) to very good (≥0.8) agreement (**Table 3**). The item from the extension addressing duration of follow up had the highest Cohens' kappa and the item addressing the understandability and relevance of the anchor the lowest.

Table 3. Inter-rater reliability coefficients

Item	Weighted κ (95% CI)
Core Instrument	

Is the patient or necessary proxy responding directly to both the patient-reported outcome and the anchor?	0.80 (0.64 to 0.95)
Is the anchor easily understandable and relevant for patients or necessary proxy?	0.70 (0.66 to 0.76)
Has the anchor shown good correlation with the patient-reported outcome measure?	0.90 (0.86 to 0.94)
Is the MID precise?	0.80 (0.67 to 0.87)
Does the threshold or difference between groups on the anchor used to estimate the MID reflect a small but important difference?	0.74 (0.71 to 0.79)
Extension for Transition Ratings	
Is the amount of elapsed time between baseline and follow-up measurement for MID estimation optimal?	0.94 (0.91 to 0.96)

CI, confidence interval; MID minimal important difference; κ, kappa

DISCUSSION

Principal findings

We have developed an instrument – the first of its kind – to evaluate the design, conduct and analysis of studies measuring anchor-based MIDs. All five criteria in the core credibility instrument proved reliable with good to excellent agreement between reviewers. The items addressing the understandability and relevance of the anchor, and whether the threshold on the anchor represents a small but important difference had lower but still very satisfactory kappa estimates.

Strengths and limitations

Strengths of our study include the use of prior literature and study team expertise in development of our criteria, and modification based both on expert feedback and extensive experience in applying the instrument. Similar methods have proved successful for developing methodological quality appraisal standards across a wide range of topics[76-80]. We undertook a rigorous assessment that demonstrated the high reliability of the instrument.

Our study has limitations. First, although a multidisciplinary team with a broad range of content and methodological expertise led the development of the credibility instrument, these individuals represent only a fraction of worldwide

experts in PRO and MID methodology. Second, given researchers in the field have not reached a consensus regarding optimal anchor-based approaches, types of anchors and analytical methods, methodological issues may subsequently emerge that will require modification of the instrument. Third, reviewers who participated in our reliability study all had graduate-level methodology training and received extensive additional instruction on MID methodology, extracted data from at least 30 studies reporting MID estimates, and participated in pilot testing with different iterations of the instrument. Thus, reliability may be lower in less well-trained and instructed individuals. We have, however, developed detailed instructions and examples included in this paper and the appended material that are likely to enhance reliability in those with less experience than the raters who participated in this study. Fourth, we were unable to assess inter-rater reliability for three items in the extension for transition rating anchors, as only 3% of studies included in our inventory of MID estimation studies evaluated the correlations necessary to judge the validity of transition rating anchors.

Implications and future research

Since the MID was first introduced in 1989[3], methods for calculating the MID have evolved. In our linked inventory of published anchor-based MIDs, we identified 17 statistical methods, each with its own merits and limitations. We also found varying quality of the anchor, and the threshold selected for defining the MID may not always be optimal. Different methodological and statistical approaches to calculate MIDs will yield different estimates for the same PROM[51,81]. Given the multiplicity of MID estimates often available for a given PROM and unstandardized methodology, researchers and decision-makers in search of MIDs need to critically evaluate the quality of the available estimates. Our credibility instrument provides a comprehensive approach to assessing the credibility of anchor-based MID estimates. Widespread adoption and implementation of our credibility instrument will not only facilitate improved appraisal of MIDs by users such as trialists,

systematic reviewers, guideline developers, clinicians, funders, and policymakers, but also guide the development of trustworthy MID estimates.

In developing our inventory of anchor-based MIDs, and in other related work[82], we found that the literature often includes a number of candidate MIDs for the same PROM. Moreover, the magnitude of these estimates sometimes varies widely. Several other researcher groups have made similar observations, stressing the importance of improved understanding of factors influencing the magnitude of MIDs[33,51,83-85]. Future research should, therefore, focus on understanding how different methodological and statistical approaches contribute to variability in MIDs.

Our instrument focuses on the methodological issues that could potentially lead to flawed and thus misleading MID estimates, which may in part explain why different methods may yield variable estimates. Variability in MIDs may, however, also be related to a multitude of other factors, including the clinical setting, patient characteristics (e.g. age, gender, disease severity, diagnosis), intervention and duration of follow-up. Findings from subsequent investigations may thus provide insights into the appropriate use – with respect to context and trustworthiness – of MIDs for interpretation of PROMs in clinical research and practice.

CONCLUSIONS

In order to better inform management choices, patients, clinicians, and researchers require knowledge of MIDs to facilitate interpretation of treatment effects on PROMs. Consideration of the credibility of an MID involves complex judgments. We have developed a reliable instrument that will allow users to distinguish between MID estimates that are more and less credible. This work not only provides guidance for addressing credibility of MIDs to optimize the presentation and interpretation of results from PROMs in clinical trials, systematic reviews health

technology assessments and clinical practice guidelines, but also has important implications for how investigators should conduct future MID estimation studies.

What is already known on this topic

- Interpreting results from patient-reported outcome measures (PROMs) is critical for optimal health care decision-making
- The minimal important difference (MID), which provides a measure of the smallest change in a PROM that patients consider important, can greatly facilitate judgments regarding magnitude of effect on PROM outcomes
- Credibility of MID estimates varies, and guidance on determining credibility has remained, until now, very limited

What this study adds

- We have developed an instrument – the first of its kind – to evaluate the design, conduct and analysis of studies measuring MIDs
- This instrument will allow users to distinguish between MID estimates that are more and less credible to optimize the presentation and interpretation of

results from PROMs in clinical trials, systematic reviews, health technology assessments and clinical practice guidelines

- This instrument will also promote higher methodologic standards for robust anchor-based MID estimation

Linked articles

Alonso Carrasco-Labra*, Tahira Devji*, Anila Qasim, Mark Phillips, Niveditha Devasenapathy, Dena Zeraatkar, Meha Bhatt, Xuejing Jin, Romina Brignardello-Petersen, Olivia Urquhart, Farid Foroutan, Stefan Schandelmaier, Hector Pardo-Hernandez, Robin WM Vernooij, Hsiaomin Huang, Yamna Rizwan, Reed Siemieniuk, Lyubov Lytvyn, Bradley C Johnston, Donald L Patrick, Shanil Ebrahim, Toshi Furukawa, Gihad Nesrallah, Holger J Schunemann, Mohit Bhandari, Lehana Thabane, Gordon H Guyatt. Minimal important difference estimates for patient-reported outcomes: The MID inventory. Submitted to *BMJ*. Dec 2018.

Contributors statement

TD, ACL, GHG, BCJ, GN, SE conceived the study idea; TD, ACL, AQ, MP, GG led the development of the credibility instrument; TD, ACL, AQ, MP, ND, DZ, MB, XJ, RBP, OU, FF, SS, HPH, RWMV, HH, YR, RAS, and LL extracted data and assessed the credibility of MIDs in our inventory for the reliability analyses; TD and ACL wrote the first draft of the manuscript; TD, ACL, GG, AQ, MP, ND, DZ, RBP, OU, SS, HPH, RWMV, LL, BCJ, DLP, SE, TF, GN, HJS, MB, LT interpreted the data analysis and critically revised the manuscript. TD and ACL are the guarantors.

Funding statement

Competing interests statement

All authors have completed the ICMJE uniform disclosure form and declare no support from any organization for the submitted work. There are no other relationships or activities that could appear to have influenced the submitted work.

Ethical approval statement: Not required.

Data sharing statement: No additional data available.

Transparency statement: TD, ACL and GHG affirm that the manuscript is an honest, accurate, and transparent account of the recommendation being reported; that no important aspects of the recommendation have been omitted; and that any discrepancies from the recommendation as planned (and, if relevant, registered) have been explained.

References

1. US Food and Drug Administration (FDA). Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. *Health and Quality of Life Outcomes.* 2006;4:1-20.
2. Guyatt GH, Juniper EF, Walter SD, Griffith LE, Goldstein RS. Interpreting treatment effects in randomised trials. *BMJ: British Medical Journal.* 1998;316(7132):690.
3. Jaeschke R, Singer J, Guyatt GH. Measurement of health status: ascertaining the minimal clinically important difference. *Controlled clinical trials.* 1989;10(4):407-415.
4. Schünemann HJ, Guyatt GH. Commentary—goodbye M (C) ID! Hello MID, where do you come from? *Health services research.* 2005;40(2):593-597.
5. Brożek JL, Guyatt GH, Schünemann HJ. How a well-grounded minimal important difference can enhance transparency of labelling claims and improve interpretation of a patient reported outcome measure. *Health and quality of life outcomes.* 2006;4(1):69.
6. McGlothlin AE, Lewis RJ. Minimal clinically important difference: defining what really matters to patients. *Jama.* 2014;312(13):1342-1343.
7. Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR. Methods to explain the clinical significance of health status measures. *Mayo Clinic proceedings.* 2002;77(4):371-383.
8. Nowell LS, Norris JM, White DE, Moules NJJIJoQM. Thematic analysis: Striving to meet the trustworthiness criteria. 2017;16(1):1609406917733847.
9. Johnston BC, Thorlund K, da Costa BR, Furukawa TA, Guyatt GH. New methods can extend the use of minimal important difference units in meta-analyses of continuous outcome measures. *J Clin Epidemiol.* 2012;65(8):817-826.
10. Johnston BC, Thorlund K, Schunemann HJ, et al. Improving the interpretation of quality of life evidence in meta-analyses: the application of minimal important difference units. *Health Qual Life Outcomes.* 2010;8:116.
11. Juniper EF, Guyatt GH, Willan A, Griffith LE. Determining a minimal important change in a disease-specific quality of life questionnaire. *Journal of clinical epidemiology.* 1994;47(1):81-87.
12. Turner D, Schunemann HJ, Griffith LE, et al. Using the entire cohort in the receiver operating characteristic analysis maximizes precision of the minimal important difference. *J Clin Epidemiol.* 2009;62(4):374-379.

13. Turner D, Schunemann HJ, Griffith LE, et al. The minimal detectable change cannot reliably replace the minimal important difference. *J Clin Epidemiol.* 2010;63(1):28-36.
14. Yalcin I, Patrick DL, Summers K, Kinchen K, Bump RC. Minimal clinically important differences in Incontinence Quality-of-Life scores in stress urinary incontinence. *Urology.* 2006;67(6):1304-1308.
15. Ebrahim S, Vercammen K, Sivanand A, et al. Minimally Important Differences in Patient or Proxy-Reported Outcome Studies Relevant to Children: A Systematic Review. *Pediatrics.* 2017;139(3).
16. Devji T, Guyatt GH, Lytvyn L, et al. Application of minimal important differences in degenerative knee disease outcomes: a systematic review and case study to inform BMJ Rapid Recommendations. *BMJ Open.* 2017;7(5).
17. De Vet HC, Terwee CB, Mokkink LB, Knol DL. Concepts, theories and models, and types of measurements. In: *Measurement in medicine: a practical guide.* Cambridge University Press; 2011.
18. Shoukri MM. *Measures of interobserver agreement and reliability.* CRC press; 2010.
19. Heppner PP, Wampold BE, Kivlighan DJCL. Research design in counseling: Research, statistics, & program evaluation. 2007.
20. Kaplan RM, Saccuzzo DP. *Psychological testing: Principles, applications, and issues.* Nelson Education; 2017.
21. Shrout PE, Fleiss JLJPb. Intraclass correlations: uses in assessing rater reliability. 1979;86(2):420.
22. Arpinelli F, Bamfi F. The FDA guidance for industry on PROs: the point of view of a pharmaceutical company. *Health Qual Life Outcomes.* 2006;4:85.
23. Barrett B, Brown D, Mundt M, Brown R. Sufficiently important difference: expanding the framework of clinical significance. *Med Decis Making.* 2005;25(3):250-261.
24. Beaton DE, Bombardier C, Katz JN, 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(2):400-405.
25. Beninato M, Portney LG. Applying concepts of responsiveness to patient management in neurologic physical therapy. *Journal of neurologic physical therapy : JNPT.* 2011;35(2):75-81.
26. Chuang-Stein C, Kirby S, Hirsch I, Atkinson G. The role of the minimum clinically important difference and its impact on designing a trial. *Pharmaceutical statistics.* 2011;10(3):250-256.
27. Copay AG, Subach BR, Glassman SD, Polly DW, Jr., Schuler TC. Understanding the minimum clinically important difference: a review

of concepts and methods. *The spine journal : official journal of the North American Spine Society.* 2007;7(5):541-546.

28. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. *J Clin Epidemiol.* 2003;56(5):395-407.

29. de Vet HC, Ostelo RW, Terwee CB, et al. Minimally important change determined by a visual method integrating an anchor-based and a distribution-based approach. *Qual Life Res.* 2007;16(1):131-142.

30. de Vet HC, Terluin B, Knol DL, et al. Three ways to quantify uncertainty in individually applied "minimally important change" values. *J Clin Epidemiol.* 2010;63(1):37-45.

31. Farivar SS, Liu H, Hays RD. Half standard deviation estimate of the minimally important difference in HRQOL scores? *Expert review of pharmacoeconomics & outcomes research.* 2004;4(5):515-523.

32. Gatchel RJ, Mayer TG, Chou R. What does/should the minimum clinically important difference measure? A reconsideration of its clinical value in evaluating efficacy of lumbar fusion surgery. *The Clinical journal of pain.* 2012;28(5):387-397.

33. Guyatt G, Schunemann H. How can quality of life researchers make their work more useful to health workers and their patients? *Qual Life Res.* 2007;16(7):1097-1105.

34. Guyatt GH, Norman GR, Juniper EF, Griffith LE. A critical look at transition ratings. *J Clin Epidemiol.* 2002;55(9):900-908.

35. Hays RD, Farivar SS, Liu H. Approaches and recommendations for estimating minimally important differences for health-related quality of life measures. *Copd.* 2005;2(1):63-67.

36. Hays RD, Woolley JM. The concept of clinically meaningful difference in health-related quality-of-life research. How meaningful is it? *PharmacoEconomics.* 2000;18(5):419-423.

37. Kamper SJ, Ostelo RW, Knol DL, Maher CG, de Vet HC, Hancock MJ. Global Perceived Effect scales provided reliable assessments of health transition in people with musculoskeletal disorders, but ratings are strongly influenced by current status. *J Clin Epidemiol.* 2010;63(7):760-766.e761.

38. Kemmler G, Giesinger J, Holzner BJJoce. Clinically relevant, statistically significant, or both? Minimal important change in the individual subject revisited. 2011;64(12):1467-1468.

39. King MT. A point of minimal important difference (MID): a critique of terminology and methods. *Expert review of pharmacoeconomics & outcomes research.* 2011;11(2):171-184.

40. Kirby S, Chuang-Stein C, Morris M. Determining a minimum clinically important difference between treatments for a patient-reported

outcome. *Journal of biopharmaceutical statistics.* 2010;20(5):1043-1054.

41. Koynova D, Luhmann R, Fischer R. A Framework for Managing the Minimal Clinically Important Difference in Clinical Trials. *Therapeutic innovation & regulatory science.* 2013;47(4):447-454.

42. Leidy NK, Wyrwich KW. Bridging the gap: using triangulation methodology to estimate minimal clinically important differences (MCIDs). *Copd.* 2005;2(1):157-165.

43. Lemieux J, Beaton DE, Hogg-Johnson S, Bordeleau LJ, Goodwin PJ. Three methods for minimally important difference: no relationship was found with the net proportion of patients improving. *J Clin Epidemiol.* 2007;60(5):448-455.

44. Molnar FJ, Man-Son-Hing M, Fergusson D. Systematic review of measures of clinical significance employed in randomized controlled trials of drugs for dementia. *J Am Geriatr Soc.* 2009;57(3):536-546.

45. Norman GR, Sridhar FG, Guyatt GH, Walter SD. Relation of distribution- and anchor-based approaches in interpretation of changes in health-related quality of life. *Med Care.* 2001;39(10):1039-1047.

46. Osoba D, Bezjak A, Brundage M, Zee B, Tu D, Pater J. Analysis and interpretation of health-related quality-of-life data from clinical trials: basic approach of The National Cancer Institute of Canada Clinical Trials Group. *European journal of cancer (Oxford, England : 1990).* 2005;41(2):280-287.

47. Rennard SI. Minimal clinically important difference, clinical perspective: an opinion. *Copd.* 2005;2(1):51-55.

48. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol.* 2008;61(2):102-109.

49. Revicki DA, Cella D, Hays RD, Sloan JA, Lenderking WR, Aaronson NK. Responsiveness and minimal important differences for patient reported outcomes. *Health Qual Life Outcomes.* 2006;4:70.

50. Sloan JA. Assessing the minimally clinically significant difference: scientific considerations, challenges and solutions. *Copd.* 2005;2(1):57-62.

51. Terwee CB, Roorda LD, Dekker J, et al. Mind the MIC: large variation among populations and methods. *J Clin Epidemiol.* 2010;63(5):524-534.

52. Tubach F, Giraudeau B, Ravaud P. The variability in minimal clinically important difference and patient acceptable symptomatic state values did not have an impact on treatment effect estimates. *J Clin Epidemiol.* 2009;62(7):725-728.

53. Wright A, Hannon J, Hegedus EJ, Kavchak AE. Clinimetrics corner: a closer look at the minimal clinically important difference (MCID). *The Journal of manual & manipulative therapy.* 2012;20(3):160-166.

54.   Wyrwich KW, Bullinger M, Aaronson N, Hays RD, Patrick DL, Symonds T. Estimating clinically significant differences in quality of life outcomes. *Qual Life Res.* 2005;14(2):285-295.

55.   Yost KJ, Eton DT. Combining distribution- and anchor-based approaches to determine minimally important differences: the FACIT experience. *Evaluation & the health professions.* 2005;28(2):172-191.

56.   Zannikos S, Lee L, Smith HE. Minimum clinically important difference and substantial clinical benefit: Does one size fit all diagnoses and patients? Paper presented at: Seminars in Spine Surgery2014.

57.   Schunemann HJ, Puhan M, Goldstein R, Jaeschke R, Guyatt GH. Measurement properties and interpretability of the Chronic respiratory disease questionnaire (CRQ). *Copd.* 2005;2(1):81-89.

58.   Deyo RA, Inui TS. Toward clinical applications of health status measures: sensitivity of scales to clinically important changes. *Health services research.* 1984;19(3):275.

59.   Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials.* 1989;10(4):407-415.

60.   Stucki G, Liang MH, Fossel AH, Katz JN. Relative responsiveness of condition-specific and generic health status measures in degenerative lumbar spinal stenosis. *Journal of clinical epidemiology.* 1995;48(11):1369-1378.

61.   Thompson M, Read J, Hutchings H, Paterson M, Harris JE. The cost effectiveness of auranofin: results of a randomized clinical trial. *The Journal of rheumatology.* 1988;15(1):35-42.

62.   Ware Jr J, Keller S. Interpreting general health measures. Quality of life and pharmacoeconomics in clinical trials. In: Philadelphia, PA: Lippincott-Raven Publishers; 1995.

63.   King M. The interpretation of scores from the EORTC quality of life questionnaire QLQ-C30. *Quality of life research.* 1996;5(6):555-567.

64.   Bergner M, Bobbitt RA, Carter WB, Gilson BS. The Sickness Impact Profile: development and final revision of a health status measure. *Medical care.* 1981:787-805.

65.   Idler EL, Angel RJ. Self-rated health and mortality in the NHANES-I Epidemiologic Follow-up Study. *American journal of public health.* 1990;80(4):446-452.

66.   Mossey JM, Shapiro E. Self-rated health: a predictor of mortality among the elderly. *American journal of public health.* 1982;72(8):800-808.

67.   Ware Jr JE, Manning Jr WG, Wells KB, Duan N, Newhouse JP. Health status and the use of outpatient mental health services. *American psychologist.* 1984;39(10):1090.

68. Brook RH, Ware Jr JE, Rogers WH, et al. Does free care improve adults' health? Results from a randomized controlled trial. *New England Journal of Medicine.* 1983;309(23):1426-1434.

69. Fayers PM, Machin D. *Quality of life: the assessment, analysis and interpretation of patient-reported outcomes.* John Wiley & Sons; 2013.

70. 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(3):207-221.

71. Guyatt GH, Jaeschke RJ. Reassessing quality-of-life instruments in the evaluation of new drugs. *PharmacoEconomics.* 1997;12(6):621-626.

72. Eton DT, Cella D, Yost KJ, et al. A combination of distribution- and anchor-based approaches determined minimally important differences (MIDs) for four endpoints in a breast cancer scale. *J Clin Epidemiol.* 2004;57(9):898-910.

73. Hagg O, Fritzell P, Nordwall A. The clinical importance of changes in outcome scores after treatment for chronic low back pain. *European spine journal : official publication of the European Spine Society, the European Spinal Deformity Society, and the European Section of the Cervical Spine Research Society.* 2003;12(1):12-20.

74. Hayes RP, Schultz EM, Naegeli AN, Curtis BH. Test-retest, responsiveness, and minimal important change of the ability to perform physical activities of daily living questionnaire in individuals with type 2 diabetes and obesity. *Diabetes technology & therapeutics.* 2012;14(12):1118-1125.

75. Schmitt J, Di Fabio RP. The validity of prospective and retrospective global change criterion measures. *Archives of physical medicine and rehabilitation.* 2005;86(12):2270-2276.

76. Akl EA, Sun X, Busse JW, et al. Specific instructions for estimating unclearly reported blinding status in randomized trials were reliable and valid. *J Clin Epidemiol.* 2012;65(3):262-267.

77. Furukawa TA, Jaeschke R, Cook D, Guyatt G. Measuring of patients' experience. In: *Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice.* New York, NY: McGraw-Hill; 2008:249-272.

78. Levine M, Ioannidis J, Haines T, Guyatt G. Harm (observational studies). In: *Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice.* New York, NY: McGraw-Hill; 2008.

79. Randolph A, Cook D, Guyatt G. Prognosis. In: *Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice.* New York, NY: McGraw-Hill; 2008.

80. Sun X, Briel M, Walter SD, Guyatt GH. Is a subgroup effect believable? Updating criteria to evaluate the credibility of subgroup analyses. *Bmj.* 2010;340:c117.

81.   Mills KA, Naylor JM, Eyles JP, Roos EM, Hunter DJ. Examining the Minimal Important Difference of Patient-reported Outcome Measures for Individuals with Knee Osteoarthritis: A Model Using the Knee Injury and Osteoarthritis Outcome Score. *J Rheumatol.* 2016;43(2):395-404.

82.   Devji T, Guyatt GH, Lytvyn L, et al. Application of minimal important differences in degenerative knee disease outcomes: a systematic review and case study to inform BMJ Rapid Recommendations. *BMJ open.* 2017;7(5):e015587.

83.   Frahm Olsen M, Bjerre E, Hansen MD, Tendal B, Hilden J, Hrobjartsson A. Minimum clinically important differences in chronic pain vary considerably by baseline pain and methodological factors: systematic review of empirical studies. *Journal of clinical epidemiology.* 2018;101:87-106.e102.

84.   Terluin B, Eekhout I, Terwee CB. The anchor-based minimal important change, based on receiver operating characteristic analysis or predictive modeling, may need to be adjusted for the proportion of improved patients. *Journal of clinical epidemiology.* 2017;83:90-100.

85.   Terluin B, Eekhout I, Terwee CB, de Vet HC. Minimal important change (MIC) based on a predictive modeling approach was more precise than MIC based on ROC analysis. *Journal of clinical epidemiology.* 2015;68(12):1388-1396.

Appendix 2. Search Strategy for Medline, January 1989 to April 2015

1.	(clinical* important difference? or clinical* meaningful difference? or clinical* meaningful improvement? or clinical* relevant mean difference? or clinical* significant change? or clinical* significant difference? or clinical* important improvement? or clinical* meaningful change? or mcid or minim* clinical* important or minim* clinical* detectable or minim* clinical* significant or minim* detectable difference? or minim* important change? or minim* important difference? or smallest real difference? or subjectively significant difference?).tw.
2.	"Quality of Life"/
3.	"outcome assessment(health care)"/or treatment outcome/or treatment failure/
4.	exp pain/
5.	exp disease attributes/or exp "signs and symptoms"/
6.	or/2–5
7.	1 and 6
8.	health status indicators/or "severity of illness index"/or sickness impact profile/or interviews as topic/or questionnaires/ or self report/
9.	Pain Measurement/
10.	patient satisfaction/or patient preference/
11.	or/8–10
12.	7 and 11
13.	limit 12 to yr="1989 -Current"
14.	(quality of life or life qualit??? or hrqol or hrql).mp.
15.	(assessment? outcome? or measure? outcome? or outcome? studies or outcome? study or outcome? assessment? or outcome? management or outcome? measure* or outcome? research or patient? outcome? or research outcome? or studies outcome? or study outcome? or therap* outcome? or treatment outcome? or treatment failure?).mp.
16.	pain????.mp.
17.	((activity or sever* or course) adj3 (disease or disabilit* or symptom*)).mp.
18.	or/14–17
19.	1 and 18
20.	(questionnaire? or instrument? or interview? or inventor* or test??? or scale? or subscale? or survey? or index?? or indices or form? or score? or measurement?).mp.
21.	(patient? rating? or subject* report? or subject* rating? or self report* or self evaluation? or self appraisal? or self assess* or self rating? or self rated).mp.
22.	(patient? report* or patient? observ* or patient? satisf*).mp.

23.	anchor base??.mp.
24.	or/20–23
25.	19 and 24
26.	limit 25 to yr="1989 -Current"
27.	13 or 26

**Appendix 3. Application of the Minimally Important Difference Credibility
Assessment Tool – Worked Examples**

Below we provide worked examples in which we have applied our instrument to
assess the credibility of two anchor-based minimal important difference (MID)
estimates, each from a published study. For each example, we first provide relevant
excerpts taken directly from the articles and highlight information critical for
informing the credibility assessment. We then provide a completed credibility
evaluation with detailed explanations supporting our judgments.

Example 1 - Escobar, A., Pérez, L. G., Herrera-Espiñeira, C., Aizpuru, F., Sarasqueta, C., De Tejada, M. G. S., ... & Bilbao, A. (2013). Total knee replacement; minimal clinically important differences and responders. *Osteoarthritis and cartilage*, *21*(12), 2006-2012.

Total knee replacement; minimal clinically important differences and responders

A. Escobar †*[a], L. García Pérez ‡[a], C. Herrera-Espiñeira §[a], F. Aizpuru ‖[a], C. Sarasqueta ¶[a], M. Gonzalez Sáenz de Tejada †[a], J.M. Quintana †#[a], A. Bilbao †[a]

† *Research Unit, Hospital Universitario Basurto, Avda. Montevideo 18, 48013 Bilbao, Spain*
‡ *Planning and Evaluation Service, Canary Islands Health Service, C/Pérez de Rozas, n°5, 4th Floor, 38004 Santa Cruz de Tenerife, Spain*
§ *Hospital Virgen de las Nieves, Avda. de las Fuerzas Armadas, 2, Granada, Spain*
‖ *Research Unit, Hospital Universitario Araba, C/Jose Atxotegi s/n, 01006 Vitoria-Gasteiz, Spain*
¶ *Research Unit, Hospital Universitario Donostia, Paseo Dr. Beguiristain s/n, 20014 San Sebastián, Donostia-San Sebastian, Spain*
# *Research Unit, Hospital of Galdakao-Usansolo, B° Labeaga s/n, Spain*

Escobar and colleagues estimate what they call minimally clinically important differences – in our terminology, minimally important difference or MIDs - for patients undergoing total knee replacement (TKR). The original publication reports approximately 40 MIDs for the pain and function domains of the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC), using two different patient cohorts. The authors use three unique anchors – rating of pain/function compared to before surgery, global satisfaction with surgical management, and a rating of whether the patient felt surgery was worthwhile. The authors used two different analytic methods – the mean change method, and Receiver Operating Characteristic (ROC) curve analysis – to estimate MIDs, and reported MIDs stratified by tertiles of baseline severity.

Below, we provide excerpts (direct quotes) from the article to perform the credibility assessment for the MID estimated for WOMAC pain using the ROC method for cohort 2.

INTRODUCTION

"The main goal of this study was to provide new data on MCID and responders at 1 year in patients who have undergone TKR, measured by pain and functional dimensions of Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) which could facilitate the interpretation of WOMAC changes."

METHODS

"The second cohort is a 1-year prospective study that took place in 15 hospitals; three in Andalusia, three in the Canary Islands and nine in the Basque Country (Spain). Consecutive patients placed on the waiting list to undergo primary TKR for osteoarthritis between September 2003 and September 2004 and between March 2005 and December 2006 and managed in any of the hospitals were eligible for the

study. We collected data from medical records and directly from patients. We sent to the patients questionnaires at baseline and 12 months post-surgery." "The data used in this study comprise a subset of patients who have completed preoperative and postoperative health related quality of life questionnaires and all the transition questions."

"We used the WOMAC that is a disease-specific, self-administered questionnaire[15]. It has a multidimensional scale made up of 24 items grouped into three dimensions: pain (five items), stiffness (two items), and physical function (17 items). We have studied pain and function dimensions through the Likert version with five response levels, representing different degrees of intensity: none (0), mild (1), moderate (2), severe (3) or extreme (4). The final scores were determined by adding the corresponding items for each dimension, and standardizing to a range of values from 0 to 100. According to recent recommendations[16] we have used the reverse option, from 0 (worst) to 100 (best). The WOMAC has been translated and validated into Spanish[17,18]."

Statistical analysis
"We used different statistical methods to calculate the cut-off values for MCID which has been defined[6] as the smallest difference between the scores in a questionnaire that the patient perceives to be beneficial. All patients had to answer two raw transition items (RTI), about their improvement or deterioration, one about pain and another about function 1 year after TKR (Compared to before surgery, how would you rate pain (functional limitation) in the same knee?). The five responses were "a great deal better", "somewhat better", "equal", "somewhat worse" and "a great deal worse". Second, we have used the Receiver Operating Characteristics (ROC) curve approach, considering the dichotomized RTI (a great deal better and somewhat better vs equal, somewhat worse and a great deal worse) as the dependent variable, and the change score for each dimension as independent. As optimal cut-off value of each dimension, the one which maximized the sum of sensitivity and specificity was considered. We draw 500 bootstrap samples[20], calculated their respective ROC curves and derived the 95% confidence interval (CI)."

"To assess the usefulness of RTI in establishing the MCID, we have evaluated their validity and reliability[12]. Validity through the association between RTI and the change score in pain, by means of partial correlation coefficients, controlling for baseline score. We hypothesized that correlation should be higher than 0.5[21]. We evaluated the correlation among RTI and pre and post-scores by Spearman's correlation coefficient."

RESULTS
Samples description

"There were 415 and 497 patients in the first and second cohorts respectively. In both groups, about 70% were females, the mean age was 71 years old and the mean Body Mass Index (BMI) was 30." "As it was expected, there were large improvements, both in pain and function, about 34 and 32 points, respectively [at 1 year]." "In comparing baseline pain, function, age, BMI and gender, … In the second [cohort], non-included patients scored five points higher in pain and function and, there were 6% more females (data not shown)."

RTI

"The partial correlation coefficients between RTI-change scores in pain [was] … 0.62 (second cohort)." "The correlation between RTI-baseline pain was … −0.05 in the … second cohort, while with the 1-year score it was … 0.47."

MCID for pain

"Table II shows data on the SEM and MCID in the pain dimension with their 95% CI along with the percentage of patients who were above those values." "The global value obtained by ROC analysis was about 22 points."

Table II
MCID data for the WOMAC pain domain

	Second cohort ($n = 497$)	
	Cut-off value (95% CI)	Patients %§
SEM*	8.3	
MCID† global‡	28.1 (25.1−31.0) $n = 207$‖	63.4
MCID† tertiles‡		
Worst	44.5 (39.9−49.2) $n = 67$‖	69.6
Medium	27.1 (23.4−30.7) $n = 71$‖	
Best	13.1 (8.8−17.4) $n = 69$‖	
ROC global	23.5 (23.1−23.8)	72.4
Worst	40.7 (40.2−41.1)	71.6
Medium	25.0 (24.7−25.3)	66.0
Best	9.0 (7.8−10.2)	75.0
Was the surgery worthwhile?		
- Patient's answer: Probably YES, n (%)	103 (20.7)	
MCID†	25.6 (21.6−29.7)	64.0
What is your global level of satisfaction with surgical management		
- Patient's answer: Somewhat satisfied, n (%)	138 (27.8)	
MCID†	27.5 (23.7−31.4)	63.4

Tertiles of pain: first cohort (second cohort): worst: ≤37.5 (≤35); medium: 38−50 (35.5−50); best: >50 (>50).
ROC: calculated as the point that maximized the sum of sensitivity and specificity.
* SEM: standard error of measurement.
† MCID: Minimal clinically important difference.
‡ Calculated as mean change in those patients who were "somewhat better".
§ Percentage of patients exceeding the cut-off value.
‖ Sample size in the "somewhat better category".

MINIMAL IMPORTANT DIFFERENCE CREDIBILITY ASSESSMENT TOOL

Is the patient or necessary proxy responding directly to both the patient-reported outcome measure and the anchor?

- ☑ Yes
- ☐ No
- ☐ Impossible to tell

If a clinician or anyone else is responding to the anchor directly and the patients are capable of providing this information, the answer should be "no." Any other <u>necessary</u> proxy (e.g. caregiver, parent, wife, relative) responding to the anchor, the answer is "yes".

Rationale: Patients completed preoperative and postoperative health related quality of life questionnaires and all the transition questions.

Is the anchor easily understandable and relevant for patients or necessary proxy?

- ☑ Definitely yes
- ☐ To a great extent
- ☐ Not so much
- ☐ Definitely no
- ☐ Impossible to tell

With "easily understandable and relevant" we mean that, when presented with the anchor (either actually presented or hypothetically) as an outcome, and without too much education, the patients are able to understand the data provided for the outcome (anchor) and use it easily for decision-making. For example, when addressing a multi-item patient-reported outcome measure addressing the potential therapeutic effects of an intervention for iron-deficiency anemia, an anchor of patient's global rating of improvement in fatigue may be easier to understand and more relevant for decision-making than serum iron levels.

If you were a patient, how would you answer this question?

Rationale: The anchor is a transition rating that asks, "Compared to before surgery, how would you rate pain (functional limitation) in the same knee?). The five responses were "a great deal better", "somewhat better", "equal", "somewhat worse" and "a great deal worse".

Has the anchor shown good correlation with the patient-reported outcome measure?

- ☐ Definitely yes (≥0.7)
- ☑ To a great extent (≥0.5 to <0.7)
- ☐ Not so much (≥0.3 to <0.5)
- ☐ Definitely no (<0.3)
- ☐ Not reported

This assessment is made using the correlation coefficients reported by the authors. If the anchor is a transition question then this is correlation between the transition item and the PROM change score. For any other anchor, this is the correlation between the change in the anchor and the change in the PROM. If the study is cross-sectional, this is the correlation between the anchor and the PROM score. Only consider the absolute value of the correlation coefficient.

Rationale: 0.62

Is the MID precise?

☑ Definitely yes (<10% **or** ≥200 patients)
☐ To a great extent (11-25% **or** 150-199 patients)
☐ Not so much (26-49% **or** 100-149 patients)
☐ Definitely no (≥50% **or** <100 patients)
☐ Impossible to tell

*Precision around the MID estimate is quantified by the width of the 95% CI and expressed as a percentage. For example, if the MID estimate is 23.5 and the 95% CI ranges from 23.1 to 23.8, then precision may be calculated as: 23.8 – 23.1 / 23.5 * 100 = 3%. According to our guide provided for our responses to this credibility question, a result of 3% would warrant a rating of definitely yes. In many cases, the authors may not report any measure of variability (e.g. SD, SE, 95% CI). In these situations, we ask that you consider the sample size used to estimate the MID. We provide ranges for both situations (i.e. percentage of the confidence interval width in relation to the MID, and sample sizes) to help inform your judgment. If the judgments according to the two criteria differ, we suggest using the higher (more permissive) of the two ratings.*
Rationale: MID estimate: 23.5; 95% CI: 23.1 to 23.8
(23.8 – 23.1) / 23.5 * 100 = 3%

Does the threshold or difference between groups on the anchor used to estimate the MID reflect a small but important difference?

☐ Definitely yes
☐ To a great extent
☑ Not so much
☐ Definitely no
☐ Impossible to tell

Establishing the degree of change on a PROM that constitutes the MID requires some knowledge about the degree of change on the anchor that is small but important to patients. In addition to inspecting the threshold on the anchor, it is necessary to judge whether the method of analysis indeed calculates a small but important difference. Below, we present examples and provide associated guidance.

For transition rating anchors, consider the wording and number of responses. For instance, the mean change in PROM score in patients with a transition rating anchor scale designation of 'a little better' on a seven-point scale including the categories 'much worse, somewhat worse, a little worse, no change, a little better, somewhat better, much better,' as reflecting an MID would warrant a definitely yes, whereas a choice of "much better" would warrant a definitely no.

In some cases, authors may use a threshold for their analysis and include only patients who achieved this threshold; other times, they may include patients who achieved this threshold or greater. For instance, the investigators may define the MID as the mean change in the PROM score in patients who achieved a ≥5% change in weight loss. This approach includes even those patients who had a 10%, 30% or 50% reduction in weight loss and thus would warrant a definitely no.
Rationale: Anchor question: "Compared to before surgery, how would you rate pain in the same knee?". Response options: "a great deal better", "somewhat better", "equal", "somewhat worse" and "a great deal worse". Groups compared: "a great deal

better" and "somewhat better" vs "equal", "somewhat worse" and a "great deal worse". We have suggested a rating of 'not so much', as there are only 2 levels representing improvement on the anchor: "somewhat better" and "a great deal better". It is possible that "somewhat better" may reflect a change in pain that is small but important; however, the limited number of categories for improvement will likely lead patients who have experienced a change that is moderate, who would not consider themselves as being "a great deal better", to rate themselves as "somewhat better", which would lead to an overestimate of the MID.

PROM patient reported outcome measure; MID minimal important difference; CI confidence interval, SD standard deviation, SE standard error

MINIMAL IMPORTANT DIFFERENCE CREDIBILITY ASSESSMENT TOOL – EXTENSION FOR TRANSITION RATINGS

Is the amount of elapsed time between baseline and follow-up measurement for MID estimation optimal? ☐ Definitely yes (≤ 1 month) ☐ To a great extent (>1 to ≤2 months) ☐ Not so much (>2 months to ≤3 months) ☑ Definitely no (>3 months) ☐ Not reported	*If there is a range of follow-up reported, consider the following when making your judgment: If the range falls over 3 categories (e.g. 3 weeks to 3 months), then select the middle category (i.e. in this example, you would select 'to a great extent'); If the range falls over 2 categories (e.g. 6 weeks to 3 months), then select the more conservative option (longest follow-up) (i.e. in this example, you would select 'not so much')* **Rationale:** Follow-up at 1-year

To answer the next 3 questions, you first need to determine if the scale of the anchor and PROM are in the same direction.
*For each question we provide 2 guides: If higher values on the anchor and PROM represent the SAME state (i.e. both represent a better or worse condition), use **Guide A**; If higher values on the anchor and PROM represent DIFFERENT states (i.e. higher scores on the PROM are worse, while higher values on the anchor are better), use **Guide B***

Does the transition item have a substantial correlation with the PROM score at follow-up? ☑ Definitely yes ☐ To a great extent ☐ Not so much ☐ Definitely no ☐ Not reported	**Guide A** *Definitely yes (>0.2)* *To a great extent (0.1 to 0.2)* *Not so much (<0.1)* *Definitely no (negative correlation)* **Rationale:** 0.47	**Guide B** *Definitely yes (<-0.2)* *To a great extent (-0.1 to -0.2)* *Not so much (>-0.1)* *Definitely no (positive correlation)*
Does the transition item correlate with the PROM score at baseline? ☑ Definitely yes ☐ To a great extent ☐ Not so much ☐ Definitely no ☐ Not reported	*Definitely yes (negative correlation)* *To a great extent (<0.1)* *Not so much (0.1 to 0.2)* *Definitely no (>0.2)* **Rationale:** -0.05	*Definitely yes (positive correlation)* *To a great extent (>-0.1)* *Not so much (-0.1 to -0.2)* *Definitely no (<-0.2)*

Is the correlation of the transition item with the PROM change score appreciably greater than the correlation of the transition item with the PRO score at follow-up?	*Definitely yes (≥0.2)*	*Definitely yes (≤-0.2)*
	To a great extent (0.1 to < 0.2)	*To a great extent (-0.1 to >-0.2)*
	Not so much (0 to <0.1)	*Not so much (0 to >-0.1)*
☐ Definitely yes	*Definitely no (<0)*	*Definitely no (>0)*
■ To a great extent	**Rationale:** Correlation of the PROM change score with the transition rating = 0.62; Correlation of the PROM post score with the transition rating = 0.47. Difference in the correlations: 0.62 – 0.47 = 0.15	
☐ Not so much		
☐ Definitely no		
☐ Not reported		

PROM patient reported outcome measure; MID minimal important difference; CI confidence interval, SD standard deviation, SE standard error

Example 2 - Ornetti, P., Dougados, M., Paternotte, S., Logeart, I., & Gossec, L. (2011). Validation of a numerical rating scale to assess functional impairment in hip and knee osteoarthritis: comparison with the WOMAC function scale. *Annals of the rheumatic diseases*, *70*(5), 740-746.

Validation of a numerical rating scale to assess functional impairment in hip and knee osteoarthritis: comparison with the WOMAC function scale

Paul Ornetti,[1–3] Maxime Dougados,[3] Simon Paternotte,[3] Isabelle Logeart,[4] Laure Gossec[3]

The objective of this study was to compare the psychometric properties of a function numerical rating scale (NRS) with the function domain of the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) and with a physician rating of patient function. Ornetti et al. also estimated minimally clinically important improvement (MCII) – in our terminology, minimally important difference (MID) – values for the two PROs in patients with knee and hip osteoarthritis (OA). The authors report 8 different MIDs, including unique MIDs for both knee and hip OA patients. The authors used two separate anchors – global state and functional status – to estimate MIDs.

Below, we provide relevant excerpts from the article to perform the credibility assessment for the MCII estimated for NRS function in knee OA patients anchored to global state.

PATIENTS AND METHODS
Study population
"Data were extracted from a previously-reported prospective study (MOVE),[16] involving outpatients with hip or knee OA, as defined by the American College of Rheumatology.[18] [19] Briefly, all patients were recruited by 399 French rheumatologists in private practice. To be included, patients had to experience pain related to OA >30 mm on a 0–100 VAS [visual analogue scale] and to require treatment with NSAIDs [non-steroidal anti-inflammatory drugs]. All patients initially visited their rheumatologist and inclusion began with the onset of NSAID treatment or with a switch from one NSAID to another." "A final visit to the same rheumatologist was scheduled 4 weeks later."

Outcome measures: Function NRS
"All patients were asked to assess their functional impairment on an 11-point NRS (patient NRS), the score ranging from 0 to 10; high scores indicate a high level of disability. The patient NRS wording was: "*What is the degree of difficulty you have experienced for the daily activities during the last 48 hours due to your (knee or hip) OA*" (online supplementary data). This PRO was assessed at baseline and after 4 weeks, without knowledge of the previous result."

Other measurements
"At the baseline visit, demographic (age, gender, body mass index) and disease data (disease duration, radiological Kellgren and Lawrence grade,[20] current symptomatic slow-acting OA drugs and NSAID intake) were collected."

"At baseline and at the final visit, all patients were asked to assess the … PROs …"

"… MCII"
"The MCII was defined as the smallest change in measurement that signifies an important improvement in patient's symptoms.[13] [15]" "All patients had to assess:
- Their degree of improvement of global state (global MCII) on a three-point Likert scale (worsened function, no change, improved function). Among the patients who improved, the degree of improvement was scored on a four-point-Likert scale (poor, fair, good, excellent).[16]"

"The global … MCII values of each function scale were calculated at the final visit …"

STATISTICAL ANALYSIS
"MCII …"
"The MCII of each function scale was defined as the 75th centile of the absolute change in score among patients whose final evaluation of response to a NSAID was improved (improvement good or excellent).[16]"

RESULTS
"In all, 881 patients with knee OA were enrolled ..." "Mean age of the patients was 66.7 ± 11.1 years, 67.7% were female and mean OA duration was 4.1 ± 5.4 years. Patients had high functional impairment patient NRS (for knee (mean 5.93 ± 1.92)."

"MCII…"
"Using MCII … questions focusing on functional impairment, 53.8% of patients with knee OA … indicated a functional improvement after treatment with NSAIDs …"

"Patients with knee OA considered their global state as improved for a change of patient NRS ≥ 2.72 (global MCII) …"

DISCUSSION

"This study which enrolled a large cohort of symptomatic patients with OA requiring treatment with NSAIDs validates a new, copyright-free instrument to assess functional impairment, the patient-reported NRS."

"The use of MCII … is of increasing interest in OA clinical research[16] [17] and in routine practice[32] to define the thresholds for monitoring response to treatment.[15]"

Table 5 minimal clinically important improvement (MCII) scores for global state and functional state in patients with knee or hip OA

	Global MCII (95% CI)
Knee OA	
Patient NRS (0–10)	−2.72 (−2.92 to −2.51)
Physician NRS (0–10)	−2.50 (−2.68 to −2.32)
WOMAC function (0–100)	−17.13 (−20.07 to −14.19)

Global MCII is defined as the smallest change in global state that signifies an important improvement in a patient's symptoms.
MCII, minimal clinically important improvement; NRS, numerical rating scale; OA, osteoarthritis;
WOMAC, Western Ontario and McMaster Universities osteoarthritis index.

MINIMAL IMPORTANT DIFFERENCE CREDIBILITY ASSESSMENT TOOL

Is the patient or necessary proxy responding directly to both the patient-reported outcome measure and the anchor?

- ☑ Yes
- ☐ No
- ☐ Impossible to tell

If a clinician or anyone else is responding to the anchor directly and the patients are capable of providing this information, the answer should be "no." Any other <u>necessary proxy</u> (e.g. caregiver, parent, wife, relative) responding to the anchor, the answer is "yes".

Rationale: All patients were asked to assess their functional impairment on an 11-point NRS (patient NRS) and their degree of improvement of global state (global MCII) on a three-point Likert scale.

Is the anchor easily understandable and relevant for patients or necessary proxy?

- ☐ Definitely yes
- ☑ To a great extent
- ☐ Not so much
- ☐ Definitely no
- ☐ Impossible to tell

With "easily understandable and relevant" we mean that, when presented with the anchor (either actually presented or hypothetically) as an outcome, and without too much education, the patients are able to understand the data provided for the outcome (anchor) and use it easily for decision-making. For example, when addressing a multi-item patient-reported outcome measure addressing the potential therapeutic effects of an intervention for iron-deficiency anemia, an anchor of patient's global rating of improvement in fatigue may be easier to understand and more relevant for decision-making than serum iron levels.

If you were a patient, how would you answer this question?

Rationale: All patients had to assess their degree of improvement of global state (global MCII) on a three-point Likert scale (worsened function, no change, improved function). Among the patients who improved, the degree of improvement was scored on a four-point-Likert scale (poor, fair, good, excellent). The exact question asked to patients was not reported, and the adjectives used to describe improvement on the anchor may be challenging to quantify in terms of relative importance of improvement.

Has the anchor shown good correlation with the patient-reported outcome measure?

☐ Definitely yes (≥0.7)

☐ To a great extent (≥0.5 to <0.7)

☐ Not so much (≥0.3 to <0.5)

☐ Definitely no (<0.3)

☒ Not reported

This assessment is made using the correlation coefficients reported by the authors. If the anchor is a transition question then this is correlation between the transition item and the PROM change score. For any other anchor, this is the correlation between the change in the anchor and the change in the PROM. If the study is cross-sectional, this is the correlation between the anchor and the PROM score. Only consider the absolute value of the correlation coefficient.

Rationale: Correlation coefficient not reported.

Is the MID precise?

☒ Definitely yes (<10% **or** ≥200 patients)

☐ To a great extent (11-25% **or** 150-199 patients)

☐ Not so much (26-49% **or** 100-149 patients)

☐ Definitely no (≥50% **or** <100 patients)

☐ Impossible to tell

*Precision around the MID estimate is quantified by the width of the 95% CI and expressed as a percentage. For example, if the MID estimate is 23.5 and the 95% CI ranges from 23.1 to 23.8, then precision may be calculated as: 23.8 – 23.1 / 23.5 * 100 = 3%. According to our guide provided for our responses to this credibility question, a result of 3% would warrant a rating of definitely yes. In many cases, the authors may not report any measure of variability (e.g. SD, SE, 95% CI). In these situations, we ask that you consider the sample size used to estimate the MID. We provide ranges for both situations (i.e. percentage of the confidence interval width in relation to the MID, and sample sizes) to help inform your judgment. If the judgments according to the two criteria differ, we suggest using the higher (more permissive) of the two ratings.*

Rationale: MID estimate: -2.72; 95% CI: -2.92 to -2.51

(-2.51 – (-2.92)) / -2.72 * 100 = 15%

Does the threshold or difference between groups on the anchor used to estimate the MID reflect a small but important difference?	*Establishing the degree of change on a PROM that constitutes the MID requires some knowledge about the degree of change on the anchor that is small but important to patients. In addition to inspecting the threshold on the anchor, it is necessary to judge whether the method of analysis indeed calculates a small but important difference. Below, we present examples and provide associated guidance.*
☐ Definitely yes ☐ To a great extent ☐ Not so much ☑ Definitely no ☐ Impossible to tell	*For transition rating anchors, consider the wording and number of responses. For instance, the mean change in PROM score in patients with a transition rating anchor scale designation of 'a little better' on a seven-point scale including the categories 'much worse, somewhat worse, a little worse, no change, a little better, somewhat better, much better,' as reflecting an MID would warrant a definitely yes, whereas a choice of "much better" would warrant a definitely no.* *In some cases, authors may use a threshold for their analysis and include only patients who achieved this threshold; other times, they may include patients who achieved this threshold or greater. For instance, the investigators may define the MID as the mean change in the PROM score in patients who achieved a ≥5% change in weight loss. This approach includes even those patients who had a 10%, 30% or 50% reduction in weight loss and thus would warrant a definitely no.* **Rationale:** The authors defined the MID of each function scale as the 75th centile of the absolute change in score among patients whose final evaluation of response to an NSAID was improved (improvement of good or excellent). First, the threshold used to define the MID (i.e. good or excellent) will very likely yield an MID estimate that is larger than a small but important improvement. Second, the choice of analytical method estimates the MCII as the 75% centile of the change scores among this group of patients, which represents the lowest score that is greater than 75% of the scores, hence further inflating the MID estimate.

PROM patient reported outcome measure; MID minimal important difference; CI confidence interval, SD standard deviation, SE standard error

MINIMAL IMPORTANT DIFFERENCE CREDIBILITY ASSESSMENT TOOL – EXTENSION FOR TRANSITION RATINGS

Is the amount of elapsed time between baseline and follow-up measurement for MID estimation optimal?
- ■ Definitely yes (≤ 1 month)
- ☐ To a great extent (>1 to ≤2 months)
- ☐ Not so much (>2 months to ≤3 months)
- ☐ Definitely no (>3 months)
- ☐ Not reported

If there is a range of follow-up reported, consider the following when making your judgment: If the range falls over 3 categories (e.g. 3 weeks to 3 months), then select the middle category (i.e. in this example, you would select 'to a great extent'); If the range falls over 2 categories (e.g. 6 weeks to 3 months), then select the more conservative option (longest follow-up) (i.e. in this example, you would select 'not so much')

Rationale: Follow-up at 4 weeks

To answer the next 3 questions, you first need to determine if the scale of the anchor and PROM are in the same direction.
*For each question we provide 2 guides: If higher values on the anchor and PROM represent the SAME state (i.e. both represent a better or worse condition), use **Guide A**; If higher values on the anchor and PROM represent DIFFERENT states (i.e. higher scores on the PROM are worse, while higher values on the anchor are better), use **Guide B***

Does the transition item have a substantial correlation with the PROM score at follow-up?
- ☐ Definitely yes
- ☐ To a great extent
- ☐ Not so much
- ☐ Definitely no
- ■ Not reported

Guide A
Definitely yes (>0.2)
To a great extent (0.1 to 0.2)
Not so much (<0.1)
Definitely no (negative correlation)

Guide B
Definitely yes (<-0.2)
To a great extent (-0.1 to -0.2)
Not so much (>-0.1)
Definitely no (positive correlation)

Rationale: Correlation coefficient not reported.

Does the transition item correlate with the PROM score at baseline?
- ☐ Definitely yes
- ☐ To a great extent
- ☐ Not so much
- ☐ Definitely no
- ■ Not reported

Definitely yes (negative correlation)
To a great extent (<0.1)
Not so much (0.1 to 0.2)
Definitely no (>0.2)

Definitely yes (positive correlation)
To a great extent (>-0.1)
Not so much (-0.1 to -0.2)
Definitely no (<-0.2)

Rationale: Correlation coefficient not reported.

Is the correlation of the transition item with the PROM change score appreciably greater than the correlation of the transition item with the PRO score at follow-up?	*Definitely yes (≥0.2)*	*Definitely yes (≤-0.2)*
☐ Definitely yes	*To a great extent (0.1 to < 0.2)*	*To a great extent (-0.1 to >-0.2)*
☐ To a great extent	*Not so much (0 to <0.1)*	*Not so much (0 to >-0.1)*
☐ Not so much	*Definitely no (<0)*	*Definitely no (>0)*
☐ Definitely no	**Rationale:** Correlation coefficient not reported.	
■ Not reported		

PROM patient reported outcome measure; MID minimal important difference; CI confidence interval, SD standard deviation, SE standard error

Chapter 3: Minimal important difference estimates for patient-reported outcomes: The MID inventory

Alonso Carrasco-Labra*, Tahira Devji*, Anila Qasim, Mark Phillips, Niveditha Devasenapathy, Dena Zeraatkar, Meha Bhatt, Xuejing Jin, Romina Brignardello-Petersen, Olivia Urquhart, Farid Foroutan, Stefan Schandelmaier, Hector Pardo-Hernandez, Robin WM Vernooij, Hsiaomin Huang, Yamna Rizwan, Reed Siemieniuk, Lyubov Lytvyn, Bradley C Johnston, Donald L Patrick, Shanil Ebrahim, Toshi Furukawa, Gihad Nesrallah, Holger J Schunemann, Mohit Bhandari, Lehana Thabane, Gordon H Guyatt

*Co-first authorship

ABSTRACT

Objectives: To develop an inventory summarizing all anchor-based minimally important difference (MID) estimates for patient-reported outcome measures (PROMs) available in the medical literature and conduct an evaluation of their credibility.

Design: Systematic review to inform the development of an MID inventory.

Data sources: We searched MEDLINE, EMBASE, PsycINFO, and the PROQOLID internal library for studies published between 1989 and April 2015.

Eligibility criteria: We included primary studies empirically calculating an anchor-based MID estimate for any PROM in adults and adolescents, irrespective of the type of anchor used.

Review methods: Pairs of reviewers independently screened and selected studies, extracted data, and evaluated the credibility of the MID estimates using a new tool.

Results: In total, 338 included studies, the majority conducted in North America (112 studies) and Europe (103 studies), reported 3,389 MID estimates for 358 PROMs. To maximize the likelihood of patients experiencing change, 91 studies determined the MID in the setting of pharmacological interventions. Of the 358 PROMs, 67% (241) were classified as disease or condition specific of which 31% related to musculoskeletal disorders. Of the MID estimates, 56% (1,885 MIDs) used a global rating of change anchor. The most common credibility issues included weak correlation (735 MIDs (21%)) or no information regarding the correlation (2,405 MIDs (71%)) between the PROM and the anchor, and imprecision in the MID estimate (2,087 MIDs (62%)).

Conclusions: A large number of MID estimates for assisting in the interpretation of PROMs exist. However, the credibility of most estimates remains limited. This MID inventory will allow more effective use of MID estimates for healthcare decision making, thus improving the interpretability of studies reporting PROMs.

INTRODUCTION

Outcomes that matter to patients have become a key focus in studies evaluating the effects of healthcare interventions. Patient-reported outcome measures (PROMs), a specific type of patient-centered outcome, can be defined as information about a patient's health condition that comes directly from the patient without interpretation by a clinician or anyone else.[1] Investigators have developed PROMs measuring constructs such as function and pain; many instruments measure a number of domains that bear on a broader construct, for instance, how dyspnea in daily life, fatigue, and emotional function affect the health-related quality of life in patients with heart and lung disease.

Although undeniably important, the difficulties with intuitive understanding of PROM reports hinder inferences regarding the magnitude of change – from trivial to very large – that patients have experienced in the constructs of interest.[2] The minimal important difference (MID), initially defined as "the smallest difference that patients perceive as beneficial and that would mandate, in the absence of troublesome side effects and excessive cost, a change in the patient's management"[3] is the most common approach to facilitating the interpretation of PROMs. An update of this definition includes the patient's perception not only of the benefits but also of harms, and the possibility of an "informed proxy" as a valid informant when the patient is incapable of providing the information.[4]

Investigators use two main strategies to determine an MID: distribution and anchor-based. Distribution-based approaches that rely on the statistical characteristics of the sample fail to incorporate the patient perspective and vary widely depending on sample characteristics.[5,6] Anchor-based approaches relate a change in a PROM to an external criterion (i.e., the anchor) that is itself interpretable, and provides meaning to the change experienced in the PROM.[7] Empirical evidence suggests that estimates from distribution-based approaches differ markedly from one another and

from anchor-based approaches and should be used only when the latter are unavailable.[8,9]

Although widely accepted, the use of anchor-based MID estimates also present challenges. Investigators must conduct searches to identify reports of MIDs and when, as is often the case, the literature includes a number of candidate MIDs, choosing the most credible is likely to prove difficult.[10-12] Therefore, to facilitate the interpretation of PROMs, and to increase our understanding of and access to MIDs, we summarized all anchor-based MID estimates for PROMs available in the medical literature, and evaluated their credibility.

METHODS

Readers can find a detailed report of the methods of our review in a previously published protocol.[13] This report adheres to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) criteria that are relevant for this type of review.[14]

Eligibility criteria

We included primary studies empirically estimating an anchor-based MID for one or more PROMs (in our terminology, the target instruments) in adolescent (≥13 to 17) or adult (≥18) populations. PROMs of interest measured health-related quality of life, functional ability, symptom severity and psychological distress and well-being.[13] Using a previously published taxonomy,[15] we classified PROMs in two main categories with two and four subcategories: 1) generic (health profiles and utility measures), and 2) specific (disease/condition, symptom, function, and population specific).

We included any reported MID estimate irrespective of the participants' condition or disease, type of intervention used in the study, or nature of the anchor. We

included reports using any MID related terminology (e.g. minimally clinically important difference, subjective significant difference, clinical important difference, minimally detectable change, etc.) and any anchor with which results on the target instrument were related, irrespective of the interpretability of that anchor.[13] This included longitudinal (e.g. global rating of change, prognosis of future events, change in disease-related outcomes) and cross-sectional (e.g. comparison to another group with a different status on the same condition or domain, preference rating) designs.[5]

We excluded systematic reviews of anchor-based MID estimation studies; abstracts from conferences; studies in which authors explicitly targeted a moderate or large important difference as opposed to an MID; MIDs estimated using a combined anchor and distribution-based approach; and estimates obtained using pooled data from multiple cohorts (i.e. different primary investigations).

Literature search

We searched Medline, EMBASE, and PsycINFO for studies published between 1989 and April 2015 (the MID concept was first described in the medical literature in 1989[3]). The search strategy, adapted to each database, included terms representing the MID concept along with terms addressing PROMs (**Appendix 4**). To complement this search, we accessed the Patient Reported Outcome and Quality of Life Instruments Database (PROQOLID)[16] internal library and retrieved additional relevant citations and reviewed reference lists from relevant reviews and eligible studies.

Study selection, data collection and analysis

Teams of two reviewers independently screened titles and abstracts for potentially eligible studies. Any studies identified as potentially relevant by either screener were selected for full text evaluation, again conducted in duplicate. Reviewers

resolved disagreement by discussion or, if needed, by consultation with a third reviewer (ACL, TD).

Prior to commencing data extraction, all reviewers received extensive training and participated in calibration exercises in which reviewers abstracted and thoroughly discussed data from up to seven studies. The unit of data extraction was the MID estimates. For each MID, we abstracted information pertaining to: the country of the study; population demographics; PROM characteristics; interventions administered in the context of the MID estimation; anchor details (i.e. type, construct(s), range of options/categories/values, threshold selected to represent a "small but important difference", specific anchor-based method); MID estimate, its associated measure of variability and direction; details regarding MID determination (e.g. number of patients informing the MID estimate, duration of follow up (if applicable), analytical (or estimation) approach, correlations between the PROM and anchor). Each pair of reviewers resolved disagreements by discussion with input from a third reviewer (ACL, TD). We used descriptive statistics such as frequencies and percentages to summarize the data.

Credibility assessment

We defined credibility as "the extent to which the design and conduct of studies measuring MIDs are likely to have protected against misleading estimates".[13] We assessed the credibility of MID estimates using an instrument developed in the context of this project; we report the development of the instrument, its characteristics and reliability elsewhere. (Submitted Dec 2018 to the BMJ). The instrument is designed for assessment of an individual MID estimate; thus, each MID estimate from a single study providing multiple estimates warrants its own credibility evaluation. The tool includes two components: 1) a core instrument with five criteria applicable to any anchor-based MID estimation, and 2) an extension of the core instrument with four criteria addressing global ratings of change – also

referred to as a transition rating – anchors. With the exception of the first item, which has a yes/no response, each item in the instrument provides a five-point adjectival scale. The range of response options for remaining items include: definitely yes; to a great extent; not so much; definitely no; impossible to tell, with wording such that a response of 'definitely yes' suggests no issues regarding the credibility of the MID estimate. Two reviewers independently conducted the credibility evaluation, resolving disagreements by discussion with input and the presence of a third reviewer for quality control (ACL, TD).

The results of this systematic review informed the development of an inventory that includes all identified anchor-based MID estimates.

RESULTS

Search Results

Of 5,656 unique citations, 1,716 proved potentially eligible after title and abstract screening, of which 338 studies were eligible after full text evaluation (**Figure 4**). For individuals in search of a specific MID, we have created a comprehensive reference list of all included studies classified according to clinical area and indexed by each PROM (**Appendix 5**).
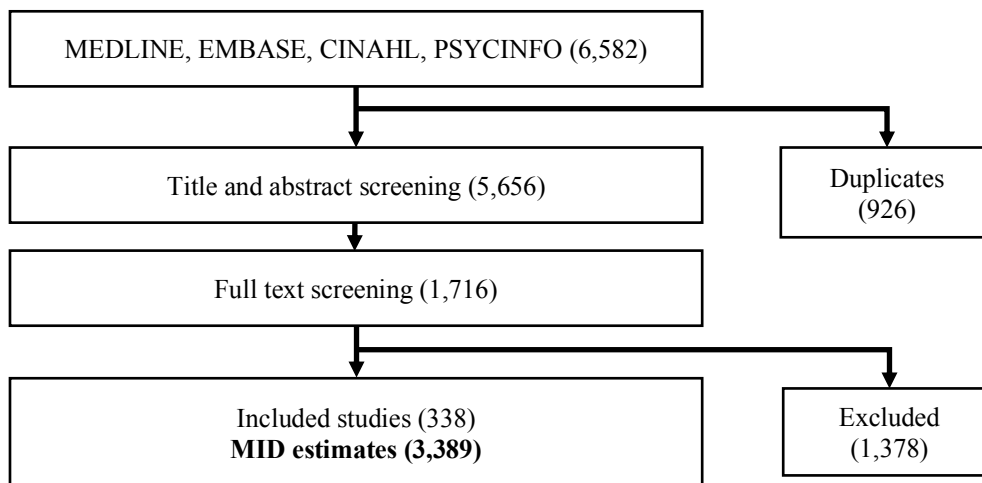
Figure 4. PRISMA flowchart for study selection process

Study level characteristics

Table 4 describes the study characteristics. Of 338 included studies, the majority were conducted in North America or Europe with the most common area of study being musculoskeletal and other pain. To maximize the likelihood of participants experiencing change, many investigators conducted their studies in the context of patients receiving interventions, most commonly pharmacological, surgical or invasive interventions, and rehabilitation. Among all studies, 44% were conducted exclusively in adults under age 65, 45% in adults of all ages, 2% exclusively in those over 65, whereas 0.5% were exclusively in adolescents or in adolescents and adults of all ages. **Figure 5a** shows that most of the studies (n=270) reported no more than two PROMs, while 60 included between three to five PROMs.
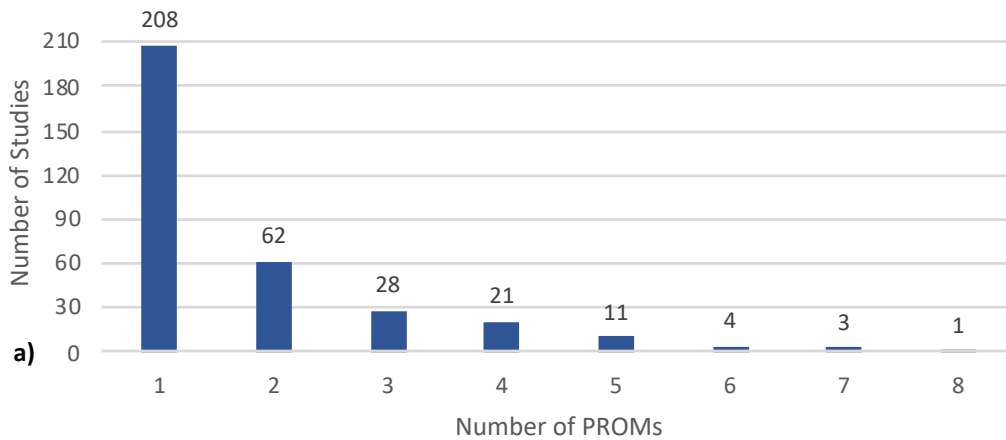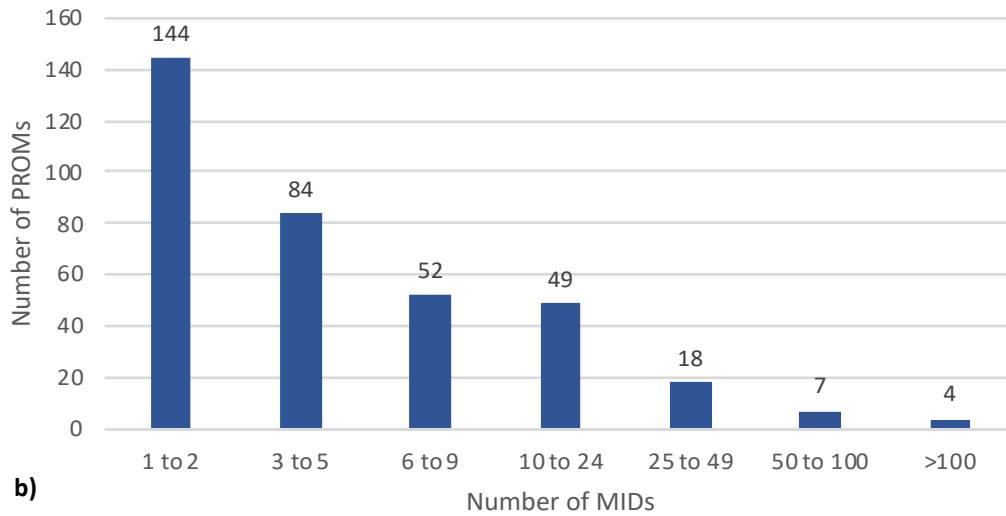
b)
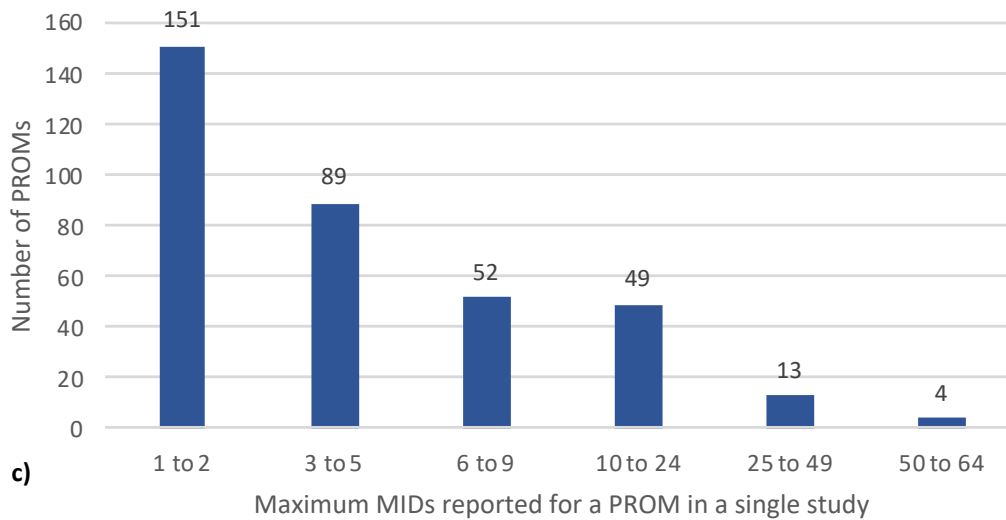
c)

Figure 5. a) Frequency of PROM reported in individual studies; **b)** Frequency of MIDs available for PROMs; **c)** Maximum number of MIDs reported for a PROM in a single study

PROM characteristics

Table 4 presents characteristics of the 358 PROMs for which MIDs were available, majority of which were specific for a disease/condition, a symptom or a function;

while only a few PROMs were classified as generic health profiles or utility indices. Disease/condition-specific PROMs most commonly addressed musculoskeletal disorders, cancer, and neurologic conditions. Symptom-specific PROMs most frequently evaluated non-specific or non-musculoskeletal pain, musculoskeletal symptoms, fatigue and dyspnea; and function-specific PROMs frequently assessed physical function and sleep. **Figure 5b** shows that most PROMs have more than one MID available, with four PROMs having more than 100 MID estimates available.

MID characteristics

Table 4 presents the characteristics of the 3,389 individual MID estimates for the 358 PROMs reported in the 338 eligible studies. Most studies addressed the MID related to participants' improvement, with relatively few studies addressing worsening of condition or conducting analyses under the assumption that MIDs on the target instrument were similar for improvement and deterioration. **Figure 5c** presents the maximum number of MIDs reported for a PROM in a single study.

Most of the MID estimates (n=305) were generated from studies using longitudinal designs, (i.e. patients provided responses to the target instrument on two occasions, along with a global rating of change or a measure of satisfaction administered at follow-up; alternatively, change in another PROM or clinical endpoint, or the occurrence of an event was evaluated at follow-up), as opposed to cross-sectional study designs (i.e. investigators either asked participants to compare their status on the target domain to others at a single point in time, or the investigators compared target instrument scores from groups that differed on the anchor).

Anchor type and anchor-based methods

The anchor type (i.e. the source of information) and anchor method (i.e. nature of anchor) varied considerably across MID estimations (**Table 4**). Investigators

typically used anchors in which patients reported their own status (2,706 MIDs, 80%). Common patient-reported anchors included the use of a transition rating, accounting for 1,756 (65%) MIDs; measures of satisfaction (233 MIDs, 9%); occurrence of an event (e.g. incontinence episodes) or other PROMs assessing health status (e.g. pain visual analogue scale, health assessment questionnaire (HAQ) disability index, Short Form-36) (441 MIDs, 16%). Investigators used a proxy as the source of information for the anchor for 356 MID estimates (11%), which was often informed by a clinician (332 MIDs, 93%) providing their impression of change in health status using a transition rating, or assessing performance status or disease activity. Investigators used other anchors such as clinical or laboratory data (e.g. hemoglobin level, number of metastatic sites, forced vital capacity), performance-based measures (e.g. accelerometry data, best-corrected visual acuity), and administrative data (e.g. occurrence of death and rehospitalization) less frequently.

Table 4. Characteristics of the included studies, PROMs and reported MIDs

	Regions: count (%)	
	North America	112 (33)
	Europe	103 (30)
	Asia	14 (4)
	Australia	11 (3)
	South America	1 (0)
	Africa	1 (0)
	Multiple continents	18 (5)
	Not reported	78 (23)
Study level data (n=338)		
	Most common interventions: count (%)	
	Pharmacological	91 (27)
	Surgical/invasive	53 (16)
	Rehabilitation	37 (11)
	No intervention	9 (3)
	Alternative medicine	5 (1)
	Behavior	2 (1)
	Other	141 (42)
	Design: count (%)	

	Longitudinal	305 (90)
	Cross-sectional	16 (5)
	Both	16 (5)
	Unclear	1 (0.3)

	Type of PROM: count (%)	
	Disease/condition specific	**241 (67)**
	Musculoskeletal disorders	75 (31)
	Cancer	43 (18)
	Neurologic	24 (10)
	Urologic/Gynecologic	17 (7)
	Upper respiratory	14 (6)
	Other	68 (28)
	Symptom specific	**64 (18)**
	Non-specific/non- Musculoskeletal pain	21 (33)
	Musculoskeletal symptoms	14 (22)
PROM level data	Fatigue	9 (14)
(n=358)	Dyspnea	8 (13)
	Other	12 (19)
	Function specific	**21 (6)**
	Physical function	11 (52)
	Sleep	4 (19)
	Sexual function	2 (10)
	Work limitations	2 (10)
	Social function	1 (5)
	Activities of daily living	1 (5)
	Utility index	**16 (4)**
	Generic health profile	**13 (4)**
	Other	**3 (0.8)**

	MID direction: count (%)	
	Improvement	2,288 (68)
	Worsening	584 (17)
	Improvement/worsening	380 (11)
	Unclear	137 (4)
	Anchor-based methods: count (%)	
	Global rating of change	1,885 (56)
	Change in disease related outcomes	777 (23)
MID level data	Comparison to another group	454 (13)
(n=3,389)	Satisfaction scale	238 (7)
	Combination of methods	23 (0.7)
	Prognosis of future events	12 (0.4)
	Anchor type: count (%)	
	Self-reported	2,706 (80)
	Proxy-reported	356 (11)
	Laboratory data	121 (4)
	Performance-based measure	76 (2)
	Combination of types	45 (1)

Self and proxy reported		22 (1)	
Administrative data		13 (0.4)	
Unclear		50 (1)	

PROM, Patient-reported outcome measure; MID, Minimal important difference estimate

Analytical approach for MID estimation

After the anchor is selected and participants are classified according to the magnitude of difference on the anchor that is small but important to patients, investigators have used a variety of analytical approaches to compute the MID estimate (**Table 5**). In longitudinal studies, investigators most frequently examined the change in the target instrument in those who experienced a small but important change on the anchor or compared to the change in another group (e.g. patients reporting no change). Less frequently, authors used a receiver operating characteristic (ROC) curve analysis, and only infrequently other approaches. In cross-sectional studies, investigators most frequently compared scores on the target instrument in groups that differed on the anchor, but also quite frequently used regression approaches.

Table 5. Analytical approach according to study design and operational definition (n=3,389)

	Analytical approach	**n (%)**	**Operational definition**
Longitudinal design (n=2,871)	Mean change	1,425 (50)	The MID is the mean change in PROM scores over time within the subgroup of participants who reported a small but important improvement (or worsening).
	Mean difference	576 (20)	The MID is the difference in PROM scores over time in the participants in one group minus the mean change in PROM scores over time in the participants in another group. The participants in the defined groups typically have a different status on the same condition or disease-related outcome. When a global rating of change anchor is used, often the participants who reported a small but important improvement (or worsening) are compared to those in the no change group.
	Receiver operating	519 (18)	The MID is the optimal cut-off point may be defined by determining the lowest overall misclassifications (e.g. point closest to 0,1

	characteristic curve		criterion, closest to the -45° tangent line, maximizing the distance to the identity line, etc.). Other approaches to ROC analysis include but are not limited to an 80% specificity rule and the use of an optimal likelihood ratio.
	Other	351 (12)	Use of a logistic or linear regression model, ANOVA, discriminant function analysis, linkage or scale-alignment
Cross-sectional design (n=481)	Mean difference	352 (73)	The MID is the difference in PROM scores between participants who rated themselves as a little bit better (or a little bit worse) compared to another participant, and participants who rated themselves as about the same as compared to another participant; or the difference in PROM scores between participants in groups with a different status on the same condition or disease-related outcome.
	Other	129 (27)	Use of a logistic or linear regression model
Unclear (n=37)		37 (100)	Insufficient information reported to determine the MID analytical method

PROM, Patient-reported outcome measure; MID, Minimal important difference estimate; ROC, Receiver operating characteristic curve

Credibility assessment of available MID estimates

Table 6 presents the distribution of credibility ratings for the MID estimates. In most cases, studies met the first criterion – patients or proxies usually responded to both the target instrument and the anchor. Investigators usually chose easily understandable anchors (second criterion), but unfortunately these easily understandable anchors frequently used a threshold or difference between groups that failed to reflect a small but important change and, sometimes, were so poorly presented that judgement was not possible (fifth criterion). Investigators typically failed on the third and fourth criteria, usually neglecting to report the correlation between the target instrument and the anchor, and not enrolling sufficient patients to ensure a precise estimate of the MID. For more than 2,000 MIDs that used a global rating of change as the anchor, very few satisfied the four additional criteria in the extension of the credibility tool. The duration of time between the first and second administration of the target PROM was excessively long in over half the MIDs (more than 3 months), and very few investigators reported correlations between the transition score and the pre and post score on the target instrument.

Table 6. Credibility assessment of MID estimates

Core Credibility Items (n=3,389) Count (%)	Definitely no	Not so much	To a great extent	Definitely yes	Impossible to tell
1. Is the patient or necessary proxy responding directly to BOTH the PROM and the anchor?	620 (18)	-	-	2,716 (80)	53 (2)
2. Is the anchor easily understandable and relevant for patients or necessary proxy?	126 (4)	178 (5)	662 (20)	2,310 (68)	113 (3)
3. Has the anchor shown good correlation with the PROM?	246 (7)	489 (14)	204 (6)	45 (1)	2,405 (71)
4. Is the MID estimate precise?	1,610 (48)	477 (14)	311 (9)	552 (16)	439 (13)
5. Does the threshold or difference between groups on the anchor used to estimate the MID reflect a small but important difference?	880 (26)	713 (21)	1,282 (38)	163 (5)	351 (10)
Extension Credibility Items (n=2,075) Count (%)					
1. Is the amount of elapsed time between baseline and follow-up measurement for MID estimation optimal?	1,103 (53)	349 (17)	184 (9)	347 (17)	92 (4)
2. Does the transition item have a substantial positive correlation with the PROM score at follow-up?	10 (0)	8 (0)	11 (0)	41 (2)	2005 (97)
3. Does the transition item correlate negatively or very weakly positively with the PROM score at baseline?	9 (0)	10 (0)	12 (0)	23 (1)	2021 (97)
4. Is the correlation of the transition item with the PROM change score appreciably greater than the correlation of the transition item with the PROM score at follow-up?	22 (1)	10 (0)	9 (0)	8 (0)	2026 (98)

PROM, patient-reported outcome measure; MID, Minimal important difference

On the basis of the results of this systematic review, we have developed an inventory of anchor-based MID estimates that will allow users to search for all available MIDs for PROMs. For each MID we have summarized information pertaining to the study design, PROM characteristics, population demographics,

intervention details, MID methodology, anchor details, and assessment of credibility. Individuals interested in accessing the inventory can do so here: www.promid.org (in development).

DISCUSSION

Main findings

This effort represents the first systematic summary of all available anchor-based MID estimates for PROMs in the medical literature. We identified 338 primary studies reporting 3,389 anchor-based MID estimates for 358 PROMs across all clinical disciplines. Disease/condition-specific PROMs have the largest representation in our inventory and studies most frequently used longitudinal designs, with self-reported global ratings of change by far the most common type of anchor. The credibility of the MID estimates varied substantially and reporting issues often limited the credibility evaluation.

A number of insights emerged from this study. First, there are a large number of MID estimates available in the literature that can be used to inform the interpretation of a great many PROMs across a wide variety of clinical areas. Second, individual studies often report a number of MIDs, usually for only one or two PROMs; for individual PROMs there are often between one to five available MID estimates. Third, investigators make use of a variety of anchor-based methodologies; however, their relative merits remain to be established. Fourth, although the majority of the estimations were informed by anchors that were easily understandable and relevant, and to which patients or proxies responded directly, most studies failed to report the correlation between the PROM and the anchor, and presented issues of imprecision. Thus, there are substantial deficiencies in the methodology of most MID assessments; very large improvements in methodology are needed.

Strengths and limitations

The first strength of our work is its scope: it is likely that our inventory includes a near-complete collection of the anchor-based MIDs in adolescents and adults reported in the peer-reviewed medical literature, with a description of salient characteristics including credibility of MID estimates. We conducted extensive screening using broad inclusive criteria at a title and abstract level, minimizing the risk of missing MID estimates due to inconsistencies in terminology. We used a piloted form that underwent iterative testing to ensure it covered all relevant characteristics and methodological aspects of MID estimation studies. We conducted extensive calibration processes, selecting and extracting data in duplicate, and implementing a quality control with a third researcher checking the collected information. In addition, in the context of the development of this inventory, we created and applied a novel instrument to assess the credibility of MID estimates. The instrument proved to have high reliability (Submitted Dec 2018 to the BMJ).

This study also has limitations. The lack of standardized reporting for MID estimation studies presented challenges when building search strategies and conducting screening at title and abstract and full-text level, leaving the possibility that our search missed some available MID estimates. It is likely, however, that only a small proportion of the available MIDs published in peer-review journals included in the most common electronic databases to which our search was limited escaped detection. To ensure completeness, future updates of this inventory may need to include grey literature, and access to other less commonly utilized sources of information. Finally, our study is comprehensive only to April 2015; we are currently in the process of identifying resources to update the search, data abstraction, and credibility assessments.

Relation to prior work

To the best of our knowledge, this is the first attempt to systematically summarize all available anchor-based MID estimates in the literature. A number of reports have provided guidance for advancing the use of MID estimates to place PROM results in context and facilitate interpretation.[17] Investigators have proposed examining the magnitude of treatment effects in relation to the MID, and also examining the proportion of patients in intervention and control groups who have achieved improvements (or deteriorations) greater than the MID – a so-called "responder analysis".[18] This approach allows the presentation of pooled effect estimates using relative (risk ratio, odds ratio) and absolute measures (risk difference, number needed to treat for benefit or harm).[19]

When conducting a meta-analysis in which studies use different PROMs measuring the same construct, authors can report mean difference in MID units, as an alternative to the standardized mean difference – a measure associated with considerable challenges in interpretability.[20] Another approach suggests the use of MIDs for the calculation of the probability for trial participants to experience a treatment effect that is greater than or at least equal to the MID.[21,22] Authors have also suggested a role for MID estimates for determining sample size calculation.[7,22,23]

Implications for research and use of MID estimates

All methods presented in the previous section rely on the assumption that a credible MID estimate is available for the PROM under evaluation. Currently, determining whether an MID estimate for a given PROM is available presents two important challenges: 1) users of MIDs need to conduct comprehensive systematic reviews to identify primary studies reporting MID estimates for the PROM of interest, and 2) as our study showed, it is likely that more than one estimate would be available, requiring decisions of which estimate(s) to use. The credibility assessment of the MID will constitute a key, if not a pre-eminent criterion for this choice.

Recent publications provide examples of practical applications of MID estimates for improving the interpretation of PROMs in the context of primary studies, systematic reviews and clinical practice guidelines.[12,21,24-26] By providing easy access to available MIDs, including ratings of their credibility, this inventory aims to close the gap between MID estimation studies and subsequent application of their MID estimates in clinical research and practice by reducing the time, effort, and likelihood of error in MID estimate selection.

Since the early 2000s, more patient-centered approaches, such as emphasizing the use of PROMs and capturing the patient perspective to inform decision making, has gained attention in the medical community.[12,27,28] To use PROM results effectively, decision-makers must be able to accurately interpret the magnitude of treatment effects. Using an anchor-based MID estimate based on the patient's perspective provides the needed interpretation that then informs the trade-off between benefits, harms, and burdens of medical interventions.[29] Our inventory of the available MID estimates will greatly facilitate use of MIDs in interpreting PROM results. Future efforts will focus on making this inventory of MID estimates easily available to key stakeholders, maintaining updated records of the latest studies published in the medical literature, and including an assessment of their credibility. This resource will serve as a repository for users and developers of MID estimates, simplifying their identification and utilization in primary and secondary research, and clinical practice guidelines.

What is already known on this topic

- The use and optimal interpretation of patient-reported outcome measures (PROMs) are essential for patient-centered clinical research and practice.

- Minimal important difference (MID) estimates facilitate the interpretation of PROMs, providing a threshold that reflects patient perspectives on what constitutes a small but important change.
- Currently, the identification and selection of MID estimates is challenging for researchers and clinicians.

What this study adds
- We have created an inventory of all available anchor-based minimal important difference estimates in the medical literature, including an evaluation of their credibility.
- There are a large number of MID estimates available that can be used to inform the interpretation of a great many PROMs across a wide variety of clinical areas.

Linked articles

Tahira Devji*, Alonso Carrasco-Labra*, Anila Qasim, Mark Phillips, Niveditha Devasenapathy, Dena Zeraatkar, Meha Bhatt, Xuejing Jin, Romina Brignardello-Petersen, Olivia Urquhart, Farid Foroutan, Stefan Schandelmaier, Hector Pardo-Hernandez, Robin WM Vernooij, Hsiaomin Huang, Yamna Rizwan, Reed Siemieniuk, Lyubov Lytvyn, Bradley C Johnston, Donald L Patrick, Shanil Ebrahim, Toshi Furukawa, Gihad Nesrallah, Holger J Schunemann, Mohit Bhandari, Lehana Thabane, Gordon H Guyatt. Development and inter-rater reliability of an instrument to evaluate the credibility of anchor-based minimal important difference estimates for patient reported outcomes. Submitted to *BMJ*. Nov 2018

Contributors statement

ACL, TD, BCJ, GN, SE, GHG conceived the study idea; ACL, TD, AQ, MP, GG created the data extraction form for the MID inventory and led the development of

the credibility instrument; TD, ACL, AQ, MP, ND, DZ, MB, XJ, RBP, OU, FF, SS, HPH, RWMV, HH, YR, RAS, and LL extracted data and assessed the credibility of MIDs in our inventory; ACL and TD wrote the first draft of the manuscript; ACL, TD, GG, AQ, MP, ND, DZ, RBP, OU, SS, HPH, RWMV, LL, BCJ, DLP, SE, TF, GN, HJS, MB, LT interpreted the data analysis and critically revised the manuscript. ACL and TD are the guarantors.

Competing interests' statement

All authors have completed the ICMJE uniform disclosure form and declare no support from any organization for the submitted work. There are no other relationships or activities that could appear to have influenced the submitted work.

Ethical approval statement: Not required.

Data sharing statement: No additional data available.

Transparency statement: ACL, TD and GHG affirm that the manuscript is an honest, accurate, and transparent account of the recommendation being reported; that no important aspects of the recommendation have been omitted; and that any discrepancies from the recommendation as planned (and, if relevant, registered) have been explained.

References

1. Group. F-NBW. In: BEST (Biomarkers, EndpointS, and other Tools) Resource. Silver Spring (MD): Food and Drug Administration (US). Available from: https://www.ncbi.nlm.nih.gov/books/NBK326791/ Co-published by National Institutes of Health (US), Bethesda (MD). 2016.
2. Glassman SD, Carreon LY. Thresholds for Health-related Quality of Life measures: reality testing. The spine journal : official journal of the North American Spine Society. 2010;10(4):328-329.
3. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Controlled clinical trials. 1989;10(4):407-415.
4. Schunemann HJ, Puhan M, Goldstein R, Jaeschke R, Guyatt GH. Measurement properties and interpretability of the Chronic respiratory disease questionnaire (CRQ). Copd. 2005;2(1):81-89.
5. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. J Clin Epidemiol. 2003;56(5):395-407.
6. Wright A, Hannon J, Hegedus EJ, Kavchak AE. Clinimetrics corner: a closer look at the minimal clinically important difference (MCID). The Journal of manual & manipulative therapy. 2012;20(3):160-166.
7. Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR. Methods to explain the clinical significance of health status measures. Mayo Clinic proceedings. 2002;77(4):371-383.
8. Puhan MA, Mador MJ, Held U, Goldstein R, Guyatt GH, Schunemann HJ. Interpretation of treatment changes in 6-minute walk distance in patients with COPD. The European respiratory journal. 2008;32(3):637-643.
9. Turner D, Schunemann HJ, Griffith LE, et al. The minimal detectable change cannot reliably replace the minimal important difference. J Clin Epidemiol. 2010;63(1):28-36.
10. Terwee CB, Roorda LD, Dekker J, et al. Mind the MIC: large variation among populations and methods. J Clin Epidemiol. 2010;63(5):524-534.
11. de Vet HC, Ostelo RW, Terwee CB, et al. Minimally important change determined by a visual method integrating an anchor-based and a distribution-based approach. Qual Life Res. 2007;16(1):131-142.
12. Devji T, Guyatt GH, Lytvyn L, et al. Application of minimal important differences in degenerative knee disease outcomes: a systematic review and case study to inform BMJ Rapid Recommendations. BMJ open. 2017;7(5):e015587.
13. Johnston BC, Ebrahim S, Carrasco-Labra A, et al. Minimally important difference estimates and methods: a protocol. BMJ Open. 2015;5(10):e007953.
14. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ (Clinical research ed). 2009;339:b2535.

15. Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. Annals of internal medicine. 1993;118(8):622-629.
16. Pinotti R. PROQOLID. Journal of the Medical Library Association : JMLA. 2016;104(1):91-92.
17. Koynova D, Luhmann R, Fischer R. A Framework for Managing the Minimal Clinically Important Difference in Clinical Trials. Therapeutic innovation & regulatory science. 2013;47(4):447-454.
18. Schunemann HJ, Akl EA, Guyatt GH. Interpreting the results of patient reported outcome measures in clinical trials: the clinician's perspective. Health and quality of life outcomes. 2006;4:62.
19. Thorlund K, Walter SD, Johnston BC, Furukawa TA, Guyatt GH. Pooling health-related quality of life outcomes in meta-analysis-a tutorial and review of methods for enhancing interpretability. Research synthesis methods. 2011;2(3):188-203.
20. Johnston BC, Thorlund K, Schunemann HJ, et al. Improving the interpretation of quality of life evidence in meta-analyses: the application of minimal important difference units. Health Qual Life Outcomes. 2010;8:116.
21. McGlothlin AE, Lewis RJ. Minimal clinically important difference: defining what really matters to patients. Jama. 2014;312(13):1342-1343.
22. Neely JG, Karni RJ, Engel SH, Fraley PL, Nussenbaum B, Paniello RC. Practical guides to understanding sample size and minimal clinically important difference (MCID). Otolaryngology--head and neck surgery : official journal of American Academy of Otolaryngology-Head and Neck Surgery. 2007;136(1):14-18.
23. Angst F, Aeschlimann A, Angst J. The minimal clinically important difference raised the significance of outcome effects above the statistical level, with methodological implications for future studies. Journal of clinical epidemiology. 2017;82:128-136.
24. Brignardello-Petersen R, Guyatt GH, Buchbinder R, et al. Knee arthroscopy versus conservative management in patients with degenerative knee disease: a systematic review. BMJ open. 2017;7(5):e016114.
25. Siemieniuk RAC, Harris IA, Agoritsas T, et al. Arthroscopic surgery for degenerative knee arthritis and meniscal tears: a clinical practice guideline. BMJ (Clinical research ed). 2017;357:j1982.
26. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. J Clin Epidemiol. 2008;61(2):102-109.
27. Puhan MA, Behnke M, Devereaux PJ, et al. Measurement of agreement on health-related quality of life changes in response to respiratory rehabilitation by patients and physicians--a prospective study. Respiratory medicine. 2004;98(12):1195-1202.
28. Black N. Patient reported outcome measures could help transform healthcare. BMJ (Clinical research ed). 2013;346:f167.

29. Brozek JL, Guyatt GH, Schunemann HJ. How a well-grounded minimal important difference can enhance transparency of labelling claims and improve interpretation of a patient reported outcome measure. Health and quality of life outcomes. 2006;4:69.

Appendix 4. Search Strategy

Database: Ovid MEDLINE(R) 1946 to Present with Daily Update Search Strategy:

1	(clinical* important difference? or clinical* meaningful difference? or clinical* meaningful improvement? or clinical* relevant mean difference? or clinical* significant change? or clinical* significant difference? or clinical* important improvement? or clinical* meaningful change? or mcid or minim* clinical* important or minim* clinical* detectable or minim* clinical* significant or minim* detectable difference? or minim* important change? or minim* important difference? or smallest real difference? or subjectively significant difference?).tw.	5231
2	"Quality of Life"/	102387
3	"outcome assessment(health care)"/ or treatment outcome/ or treatment failure/	602632
4	exp pain/	281620
5	exp disease attributes/ or exp "signs and symptoms"/	2141451
6	or/2-5	2666010
7	1 and 6	2720
8	health status indicators/ or "severity of illness index"/ or sickness impact profile/ or interviews as topic/ or questionnaires/ or self report/	441974
9	Pain Measurement/	53721
10	patient satisfaction/ or patient preference/	54114
11	or/8-10	521840
12	7 and 11	1182
13	limit 12 to yr="1989 -Current"	1180
14	(quality of life or life qualit??? or hrqol or hrql).mp.	157316
15	(assessment? outcome? or measure? outcome? or outcome? studies or outcome? study or outcome? assessment? or outcome? management or outcome? measure* or outcome? research or patient? outcome? or research outcome? or studies outcome? or study outcome? or therap* outcome? or treatment outcome? or treatment failure?).mp.	737085
16	pain????.mp.	442420
17	((activity or sever* or course) adj3 (disease or disabilit* or symptom*)).mp.	150384
18	or/14-17	1335860
19	1 and 18	2758

20	(questionnaire? or instrument? or interview? or inventor* or test??? or scale? or subscale? or survey? or index?? or indices or form? or score? or measurement?).mp.	4884173
21	(patient? rating? or subject* report? or subject* rating? or self report* or self evaluation? or self appraisal? or self assess* or self rating? or self rated).mp.	96219
22	(patient? report* or patient? observ* or patient? satisf*).mp.	86741
23	anchor base??.mp.	185
24	or/20-23	4931748
25	19 and 24	2274
26	limit 25 to yr="1989 -Current"	2256
27	13 or 26	2301

Database: Embase - 1980 to April 2015 - Search Strategy:

1	(clinical* important difference? or clinical* meaningful difference? or clinical* meaningful improvement? or clinical* relevant mean difference? or clinical* significant change? or clinical* significant difference? or clinical* important improvement? or clinical* meaningful change? or mcid or minim* clinical* important or minim* clinical* detectable or minim* clinical* significant or minim* detectable difference? or minim* important change? or minim* important difference? or smallest real difference? or subjectively significant difference?).tw.	7353
2	"Quality of Life"/	202827
3	quality adjusted life year/	9537
4	exp treatment outcome/	867722
5	exp pain/	686857
6	exp disease course/	1733745
7	symptom/	83652
8	exp disease activity/	212815
9	exp disease severity/	991780
10	or/2-9	3195376
11	1 and 10	4192
12	health survey/	139499
13	exp questionnaire/	332570
14	exp interview/	119874
15	pain assessment/	59009
16	exp "named inventories, questionnaires and rating scales"/	123421
17	rating scale/	74530
18	self evaluation/	17604
19	patient satisfaction/	74674
20	or/12-19	798335

21	20 and 11	1882
22	limit 21 to yr="1989 -Current"	1880
23	(quality of life or life qualit??? or hrqol or hrql).mp.	252623
24	(assessment? outcome? or measure? outcome? or outcome? studies or outcome? study or outcome? assessment? or outcome? management or outcome? measure* or outcome? research or patient? outcome? or research outcome? or studies outcome? or study outcome? or therap* outcome? or treatment outcome? or treatment failure?).mp.	1018285
25	pain????.mp.	707707
26	((activity or sever* or course) adj3 (disease or disabilit* or symptom*)).mp.	683760
27	(bothersomeness or ((level? or degree?) adj3 bother*)).mp.	566
28	or/23-27	2283571
29	1 and 28	4127
30	(questionnaire? or instrument? or interview? or inventor* or test??? or scale? or subscale? or survey? or index?? or indices or form? or score? or measurement?).mp.	6260382
31	(patient? rating? or subject* report? or subject* rating? or self report* or self evaluation? or self appraisal? or self assess* or self rating? or self rated).mp.	134646
32	(patient? report* or patient? observ* or patient? satisf*).mp.	119788
33	anchor base??.mp.	280
34	or/30-33	6335006
35	29 and 34	3411
36	limit 35 to yr="1989 -Current"	3388
37	22 or 36	3468

Database: PsycINFO - 1967 to April 2015 - Search Strategy:

1	(clinical* important difference? or clinical* meaningful difference? or clinical* meaningful improvement? or clinical* relevant mean difference? or clinical* significant change? or clinical* significant difference? or clinical* important improvement? or clinical* meaningful change? or mcid or minim* clinical* important or minim* clinical* detectable or minim* clinical* significant or minim* detectable difference? or minim* important change? or minim* important difference? or smallest real difference? or subjectively significant difference?).tw.	1090
2	"Quality of Life"/	22076
3	Well being/	20059
4	exp Treatment Outcomes/	25350
5	exp pain/	35913

6	Symptoms/	37745
7	Disease Course/	8253
8	or/2-7	140385
9	1 and 8	413
10	Treatment Effectiveness Evaluation/	13605
11	Self-Evaluation/	7307
12	Self Report/	11334
13	Pain Measurement/	942
14	exp Questionnaires/	12287
15	exp Rating Scales/	16649
16	Client Satisfaction/	3565
17	or/10-16	62945
18	9 and 17	84
19	(quality of life or life qualit??? or hrqol or hrql).mp.	37410
20	(assessment? outcome? or measure? outcome? or outcome? studies or outcome? study or outcome? assessment? or outcome? management or outcome? measure* or outcome? research or patient? outcome? or research outcome? or studies outcome? or study outcome? or therap* outcome? or treatment outcome? or treatment failure?).mp.	59226
21	pain????.mp.	67462
22	((activity or sever* or course) adj3 (disease or disabilit* or symptom*)).mp.	33374
23	or/19-22	180640
24	1 and 23	540
25	(questionnaire? or instrument? or interview? or inventor* or test??? or scale? or subscale? or survey? or index?? or indices or form? or score? or measurement?).mp.	1296323
26	(client? rating? or subject* report? or subject* rating? or self report* or self evaluation? or self appraisal? or self assess* or self rating? or self rated).tw.	90635
27	(client? report* or client? observ* or client? satisf*).tw.	1771
28	anchor base??.mp.	51
29	or/25-28	1314342
30	24 and 29	486
31	18 or 30	500
32	limit 31 to yr="1989 -Current"	498

Appendix 5. Complete reference list of all included MID estimation studies categorized by clinical topic area

Allergy Medicine

Rhinoconjunctivitis Quality of Life Questionnaire (RQLQ)

Turner DS, Holger J.; Griffith, Lauren E.; Beaton, Dorcas E.; Griffiths, Anne M.; Critch, Jeffrey N.; Guyatt, Gordon H. Using the entire cohort in the receiver operating characteristic analysis maximizes precision of the minimal important difference. J Clin Epidemiol. 2009;62(4):374-379.

Juniper EF, Guyatt GH, Griffith LE, Ferrie PJ. Interpretation of rhinoconjunctivitis quality of life questionnaire data. The Journal of allergy and clinical immunology. 1996;98(4):843-845.

Rhinoconjunctivitis Quality of Life Questionnaire (RQLQ) - Japanese version

Higaki TO, M.; Kariya, S.; Fujiwara, T.; Haruna, T.; Hirai, H.; Murai, A.; Gotoh, M.; Okubo, K.; Yonekura, S.; Okamoto, Y.; Nishizaki, K. Determining minimal clinically important differences in Japanese cedar/cypress pollinosis patients. Allergology International. 2013;62(4):487-493.

Rhinoconjunctivitis Total Symptom Score (RTSS)

Devillier PC, O.; Vicaut, E.; De Beaumont, O.; Robin, B.; Dreyfus, J. F.; Bousquet, P. J. The minimally important difference in the Rhinoconjunctivitis Total Symptom Score in grass-pollen-induced allergic rhinoconjunctivitis. Allergy: European Journal of Allergy and Clinical Immunology. 2014;69(12):1689-1695.

Total nasal symptom score (TNSS) 5 item

Higaki TO, M.; Kariya, S.; Fujiwara, T.; Haruna, T.; Hirai, H.; Murai, A.; Gotoh, M.; Okubo, K.; Yonekura, S.; Okamoto, Y.; Nishizaki, K. Determining minimal clinically important differences in Japanese cedar/cypress pollinosis patients. Allergology International. 2013;62(4):487-493.

Allergy, Ear Nose and Throat

Rhinitis Control Assessment Test (RCAT)

Meltzer EOS, M.; Nathan, R.; Garris, C.; Stanford, R. H.; Kosinski, M. Reliability, validity, and responsiveness of the Rhinitis Control Assessment Test in patients with rhinitis. Journal of Allergy and Clinical Immunology. 2013;131(2):379-386.

Cardiology

Atrial Fibrillation Effect On QualiTy-Of-Life (AFEQT)

Dorian PB, C.; Mullin, C. M.; Bubien, R.; Godejohn, D.; Reynolds, M. R.; Lakkireddy, D. R.; Wimmer, A. P.; Bhandari, A.; Spertus, J. Interpreting changes in quality of life in atrial fibrillation: how much change is meaningful? Am Heart J. 2013;166(2):381-387.e388.

Cambridge Pulmonary Hypertension Outcome Review (CAMPHOR) Utility Index

Meads DMM, S. P.; Doughty, N.; Das, C.; Gin-Sing, W.; Langley, J.; Pepke-Zaba, J. The responsiveness and validity of the CAMPHOR Utility Index. Eur Respir J. 2008;32(6):1513-1519.

Chronic Heart Failure Questionnaire (CHQ)

Bennett SJO, Neil B.; Eckert, George J.; Embree, Jennifer L.; Browning, Sherry; Hou, Nan; Chui, Michelle; Deer, Melissa; Murray, Michael D. Comparison of quality of life measures in heart failure. Nurs Res. 2003;52(4):207-216.

Chronic Respiratory Disease Questionnaire (CRQ) / Chronic Heart Failure Questionnaire (CHQ)

Jaeschke RS, J.; Guyatt, G. H. Measurement of health status. Ascertaining the minimal clinically important difference. Control Clin Trials. 1989;10(4):407-415.

EuroQol-5D Utility Index (EQ-5D)

Walters SJB, John E. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. Qual Life Res. 2005;14(6):1523-1532.

Intensity of average breathlessness Numerical Rating Scale (NRS)

Oxberry SGB, J. M.; Clark, A. L.; Cleland, J. G. F.; Johnson, M. J. Minimally clinically important difference in chronic breathlessness: Every little helps. Am Heart J. 2012;164(2):229-235.

Intensity of worst breathlessness Numerical Rating Scale (NRS)

Oxberry SGB, J. M.; Clark, A. L.; Cleland, J. G. F.; Johnson, M. J. Minimally clinically important difference in chronic breathlessness: Every little helps. Am Heart J. 2012;164(2):229-235.

mBorg scale-rated average breathlessness intensity

Oxberry SGB, J. M.; Clark, A. L.; Cleland, J. G. F.; Johnson, M. J. Minimally clinically important difference in chronic breathlessness: Every little helps. Am Heart J. 2012;164(2):229-235.

mBorg scale-rated worst breathlessness intensity

Oxberry SGB, J. M.; Clark, A. L.; Cleland, J. G. F.; Johnson, M. J. Minimally clinically important difference in chronic breathlessness: Every little helps. Am Heart J. 2012;164(2):229-235.

Minnesota Living With Heart Failure Questionnaire (MLHF)

Bennett SJO, Neil B.; Eckert, George J.; Embree, Jennifer L.; Browning, Sherry; Hou, Nan; Chui, Michelle; Deer, Melissa; Murray, Michael D. Comparison of quality of life measures in heart failure. Nurs Res. 2003;52(4):207-216.

Short Form Health Survey 12-Item (SF-12)

Bennett SJO, Neil B.; Eckert, George J.; Embree, Jennifer L.; Browning, Sherry; Hou, Nan; Chui, Michelle; Deer, Melissa; Murray, Michael D. Comparison of quality of life measures in heart failure. Nurs Res. 2003;52(4):207-216.

Short-Form Six-Dimension (SF-6D)

Walters SJB, John E. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. Qual Life Res. 2005;14(6):1523-1532.

Dentistry

Condition-specific Oral Impacts on Daily Performances index (CS-OIDP)

Tsakos GB, Eduardo; D'Aiuto, Francesco; Pikhart, Hynek; Tonetti, Maurizio; Sheiham, Aubrey; Donos, Nikolaos. Assessing the minimally important difference in the oral impact on daily performances index in patients treated for periodontitis. J Clin Periodontol. 2010;37(10):903-909.

Dentine Hypersensitivity Experience Questionnaire (DHEQ)

Baker SRG, B. J.; Sufi, F.; Barlow, A.; Robinson, P. G. The Dentine Hypersensitivity Experience Questionnaire: a longitudinal validation study. J Clin Periodontol. 2014;41(1):52-59.

General Oral Health Assessment Index (GOHAI)

Jonsson BO, K. Evaluation of the effect of non-surgical periodontal treatment on oral health-related quality of life: estimation of minimal important differences 1 year after treatment. J Clin Periodontol. 2014;41(3):275-282.

Oral health impact profile (OHIP-14)

Locker DJ, Aleksandra; Clarke, Martha. Assessing the responsiveness of measures of oral health-related quality of life. Community Dent Oral Epidemiol. 2004;32(1):10-18.

Oral Health Impact Profile (OHIP-20)

Allen PFOS, Maeve; Locker, David. Determining the minimally important difference for the Oral Health Impact Profile-20. Eur J Oral Sci. 2009;117(2):129-134.

Oral Health Impact Profile (OHIP-G) - German population adaptation

John MTR, Daniel R.; Szentpetery, Andras; Steele, James. An approach to define clinical significance in prosthodontics. J Prosthodont. 2009;18(5):455-460.

Oral Impacts on Daily Performances index (OIDP)

Tsakos GB, Eduardo; D'Aiuto, Francesco; Pikhart, Hynek; Tonetti, Maurizio; Sheiham, Aubrey; Donos, Nikolaos. Assessing the minimally important difference in the oral impact on daily performances index in patients treated for periodontitis. J Clin Periodontol. 2010;37(10):903-909.

UK oral health-related quality-of-life measure (OHQoL-UK)

Jonsson BO, K. Evaluation of the effect of non-surgical periodontal treatment on oral health-related quality of life: estimation of minimal important differences 1 year after treatment. J Clin Periodontol. 2014;41(3):275-282.

Xerostomia Inventory (XI)

Thomson WM. Measuring change in dry-mouth symptoms over time using the Xerostomia Inventory. Gerodontology. 2007;24(1):30-35.

Dermatology

Dermatology Life Quality Index (DLQI)

Shikiar RH, Gale; Leahy, Michael; Lennox, Richard D. Minimal important difference (MID) of the Dermatology Life Quality Index (DLQI): results from patients with chronic idiopathic urticaria. Health Qual Life Outcomes. 2005;3:36.

Shikiar RW, Mary Kaye; Okun, Martin M.; Thompson, Christine S.; Revicki, Dennis A. The validity and responsiveness of three quality of life measures in the assessment of psoriasis patients: results of a phase II study. Health Qual Life Outcomes. 2006;4:71.

EuroQol-5D Utility Index (EQ-5D)

Shikiar RW, Mary Kaye; Okun, Martin M.; Thompson, Christine S.; Revicki, Dennis A. The validity and responsiveness of three quality of life measures in the assessment of psoriasis patients: results of a phase II study. Health Qual Life Outcomes. 2006;4:71.

Walters SJB, John E. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. Qual Life Res. 2005;14(6):1523-1532.

EuroQol-5D visual analogue scale (EQ-5D-VAS)

Shikiar RW, Mary Kaye; Okun, Martin M.; Thompson, Christine S.; Revicki, Dennis A. The validity and responsiveness of three quality of life measures in the assessment of psoriasis patients: results of a phase II study. Health Qual Life Outcomes. 2006;4:71.

Psoriasis Symptom Diary (PSD)

Strober BEN, J.; Mallya, U. G.; Guettner, A.; Papavassilis, C.; Gottlieb, A. B.; Elewski, B. E.; Turner-Bowker, D. M.; Shields, A. L.; Gwaltney, C. J.; Lebwohl, M. Item-level psychometric

properties for a new patient-reported psoriasis symptom diary. Value Health. 2013;16(6):1014-1022.

Self-Assessed Simplified Psoriasis Index (saSPI)

Chularojanamontri LG, C. E.; Chalmers, R. J. Responsiveness to change and interpretability of the simplified psoriasis index. Journal of Investigative Dermatology. 2014;134(2):351-358.

Short Form Health Survey 36-Item (SF-36)

Shikiar RW, Mary Kaye; Okun, Martin M.; Thompson, Christine S.; Revicki, Dennis A. The validity and responsiveness of three quality of life measures in the assessment of psoriasis patients: results of a phase II study. Health Qual Life Outcomes. 2006;4:71.

Short-Form Six-Dimension (SF-6D)

Walters SJB, John E. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. Qual Life Res. 2005;14(6):1523-1532.

Ear Nose and Throat

Annoyance visual analogue scale (VAS)

Adamchic IL, B.; Hauptmann, C.; Tass, P. A. Psychometric evaluation of visual analog scale for the assessment of chronic tinnitus. Am J Audiol. 2012;21(2):215-225.

Dizziness Handicap Inventory (DHI) - Norwegian version

Tamber A-LW, Kjersti T.; Strand, Liv Inger. Measurement properties of the Dizziness Handicap Inventory by cross-sectional and longitudinal designs. Health Qual Life Outcomes. 2009;7:101.

Loudness visual analogue scale (VAS)

Adamchic IL, B.; Hauptmann, C.; Tass, P. A. Psychometric evaluation of visual analog scale for the assessment of chronic tinnitus. Am J Audiol. 2012;21(2):215-225.

Modified Sino-Nasal Outcome Test-16 (SNOT-16)

Garbutt JS, Edward; Piccirillo, Jay. Use of the modified SNOT-16 in primary care patients with clinically diagnosed acute rhinosinusitis. Arch Otolaryngol Head Neck Surg. 2011;137(8):792-797.

Sino-Nasal Outcome Test-22 (SNOT-22)

Browne JPvdM, Jan H.; Lewsey, James D.; Lamping, Donna L.; Black, Nick. Mathematical coupling may account for the association between baseline severity and minimally important difference values. J Clin Epidemiol. 2010;63(8):865-874.

Hopkins CG, S.; Slack, R.; Lund, V. J.; Browne, J. P. Psychometric validity of the 22-item Sinonasal Outcome Test. Clin Otolaryngol. 2009;34(5):447-454.

Sino-Nasal Outcome Test-22 (SNOT-22) - Lithuanian version

Vaitkus SP, E.; Balsevicius, T.; Siupsinskiene, N.; Staikuniene, J.; Ryskiene, S.; Lisauskaite, L.; Vaitkus, J. Translation, cross-cultural adaptation, and validation of the sino-nasal outcome test (SNOT)-22 for Lithuanian patients. Eur Arch Otorhinolaryngol. 2013;270(6):1843-1848.

Ear Nose and Throat, Infectious Disease

Activity Impairment Assessment (AIA)

Quadri NL, A.; Keating, K. N.; Nafees, B.; Piccirillo, J.; Wild, D. Psychometric evaluation of the Sinonasal Outcome Test-16 and activity impairment assessment in acute bacterial sinusitis. Otolaryngol Head Neck Surg. 2013;149(1):161-167.

Sino-Nasal Outcome Test-16 (SNOT-16)

Quadri NL, A.; Keating, K. N.; Nafees, B.; Piccirillo, J.; Wild, D. Psychometric evaluation of the Sinonasal Outcome Test-16 and activity impairment assessment in acute bacterial sinusitis. Otolaryngol Head Neck Surg. 2013;149(1):161-167.

Ear Nose and Throat, Rheumatology

Oral health impact profile (OHIP-14) - Turkish version

Hayran OM, G.; Inanc, N.; Ergun, T.; Direskeneli, H. Assessment of minimal clinically important improvement by using Oral Health Impact Profile-14 in Behcet's disease. Clin Exp Rheumatol. 2009;27(2 Suppl 53):S79-84.

Emergency Medicine

Average pain numerical rating scale (NRS)

Mehling WEG, Viranjini; Acree, Michael; Pressman, Alice; Carey, Tim; Goldberg, Harley; Hecht, Frederick M.; Avins, Andrew L. Acute low back pain and primary care: how to define recovery and chronification? Spine. 2011;36(26):2316-2323.

Nausea visual analogue scale (VAS)

Hendey GWD, Neil F.; Fuller, Kimberly. Clinically significant changes in nausea as measured on a visual analog scale. Ann Emerg Med. 2005;45(1):77-81.

Pain intensity numerical rating scale (NRS)

de Vet HCWO, Raymond W. J. G.; Terwee, Caroline B.; van der Roer, Nicole; Knol, Dirk L.; Beckerman, Heleen; Boers, Maarten; Bouter, Lex M. Minimally important change determined by a visual method integrating an anchor-based and a distribution-based approach. Qual Life Res. 2007;16(1):131-142.

Kendrick DBS, Tania D. The minimum clinically significant difference in patient-assigned numeric scores for pain. Am J Emerg Med. 2005;23(7):828-832.

Bernstein SLB, Polly E.; Gallagher, E. John. Relationship between intensity and relief in patients with acute severe pain. Am J Emerg Med. 2006;24(2):162-166.

Pain numerical rating scale (NRS)

Bijur PEC, Andrew K.; Esses, David; Gallagher, E. John. Identifying the minimum clinically significant difference in acute pain in the elderly. Ann Emerg Med. 2010;56(5):517-521.e511.

Bijur PEC, Andrew K.; Esses, David; Gallagher, E. John. Problems with measurement of the minimum clinically significant difference in acute pain in elders. Acad Emerg Med. 2011;18(2):135-139.

Bijur PEL, Clarke T.; Gallagher, E. John. Validation of a verbally administered numerical rating scale of acute pain for use in the emergency department. Acad Emerg Med. 2003;10(4):390-392.

Pain severity visual analogue scale (VAS)

Gallagher EJB, Polly E.; Latimer, Clarke; Silver, Wendy. Reliability and validity of a visual analog scale for acute abdominal pain in the ED. Am J Emerg Med. 2002;20(4):287-290.

Pain visual analogue scale (VAS)

Todd KHF, K. G.; Funk, J. P.; Bonacci, R. Clinical significance of reported changes in pain severity. Ann Emerg Med. 1996;27(4):485-489.

Lee JSH, Elisabeth; Stiell, Ian G.; Wells, George A. Clinically important change in the visual analog scale after adequate pain control. Acad Emerg Med. 2003;10(10):1128-1130.

Bird SBD, E. W. Clinically significant changes in pain along the visual analog scale. Ann Emerg Med. 2001;38(6):639-643.

Kelly AM. Does the clinically significant difference in visual analog scale pain scores vary with gender, age, or cause of pain? Acad Emerg Med. 1998;5(11):1086-1090.

Kelly AM. The minimum clinically significant difference in visual analogue scale pain score does not differ with severity of pain. Emerg Med J. 2001;18(3):205-207.

Gallagher EJL, M.; Bijur, P. E. Prospective validation of clinically important changes in pain severity measured on a visual analog scale. Ann Emerg Med. 2001;38(6):633-638.

Meek RK, Anne-Maree; Hu, Xue Feng. Use of the visual analog scale to rate and monitor severity of nausea in the emergency department. Acad Emerg Med. 2009;16(12):1304-1310.

Bijur PEL, Clarke T.; Gallagher, E. John. Validation of a verbally administered numerical rating scale of acute pain for use in the emergency department. Acad Emerg Med. 2003;10(4):390-392.

Mark MSMA, T. T. S.; Choi, Y. F.; Wong, T. W. The minimum clinically significant difference in visual analogue scale pain score in a local emergency setting. Hong Kong Journal of Emergency Medicine. 2009;16(4):233-236.

Pictoral Representation of Pain (PRP)

Eken CA, O.; Bektas, F.; Bilge, U.; Kucukyilmaz, O.; Korkmaz, N.; Ankun, G.; Cete, Y. Reliability and validity of a new pain measurement tool: Pictorial representation of pain. Turkiye Acil Tip Dergisi. 2011;11(1):3-8.

Roland-Morris Disability Questionnaire (RMDQ)

Mehling WEG, Viranjini; Acree, Michael; Pressman, Alice; Carey, Tim; Goldberg, Harley; Hecht, Frederick M.; Avins, Andrew L. Acute low back pain and primary care: how to define recovery and chronification? Spine. 2011;36(26):2316-2323.

Roland-Morris Disability Questionnaire 5-item (RMDQ-5)

Friedman BWS, C. B.; Mulvey, L.; Esses, D.; Bijur, P. E.; John Gallagher, E. Derivation of an abbreviated instrument for use in emergency department low back pain research: the five-item Roland Morris Questionnaire. Acad Emerg Med. 2013;20(10):1013-1021.

Satisfaction visual analogue scale (VAS)

Singer AJT, H. C., Jr. Determination of the minimal clinically significant difference on a patient visual analog satisfaction scale. Acad Emerg Med. 1998;5(10):1007-1011.

Worst pain numerical rating scale (NRS)

Mehling WEG, Viranjini; Acree, Michael; Pressman, Alice; Carey, Tim; Goldberg, Harley; Hecht, Frederick M.; Avins, Andrew L. Acute low back pain and primary care: how to define recovery and chronification? Spine. 2011;36(26):2316-2323.

Endocrinology

Ability to Perform Physical Activities of Daily Living (APPADL)

Hayes RPS, E. M.; Naegeli, A. N.; Curtis, B. H. Test-retest, responsiveness, and minimal important change of the ability to perform physical activities of daily living questionnaire in individuals with type 2 diabetes and obesity. Diabetes Technol Ther. 2012;14(12):1118-1125.

Diabetes Medication Satisfaction (DiabMedSat)

Brod MC, Torsten; Kongso, Jens Harald; Bushnell, Donald M. Examining and interpreting responsiveness of the Diabetes Medication Satisfaction measure. J Med Econ. 2009;12(4):309-316.

Hypoglycaemia Fear Survey (HFS-II)

Stargardt TG-F, L.; Krobot, K. J.; Alexander, C. M. Fear of hypoglycaemia: Defining a minimum clinically important difference in patients with type 2 diabetes. Health and Quality of Life Outcomes. 2009;7(91).

Short Osteoporosis Quality Of Life Questionnaire (ECOS-16)

Badia XD-P, Adolfo; Lahoz, Raquel; Lizan, Luis; Nogues, Xavier; Iborra, Jordi. The ECOS-16 questionnaire for the evaluation of health related quality of life in post-menopausal women with osteoporosis. Health Qual Life Outcomes. 2004;2:41.

Well-Being Questionnaire (W-BQ28)

Speight JK, L. A.; Davies, M. J. Generic and diabetes-specific well-being in the AT.LANTUS Follow-on study: Further psychometric validation of the W-BQ28 indicates its utility in research and clinical practice in Type 2 diabetes in the UK. Diabet Med. 2012;29(9):e345-e353.

Gastroenterology

Abdominal pain numerical rating scale (NRS)

Spiegel BB, R.; Harris, L. A.; Lucak, S.; Naliboff, B.; Esrailian, E.; Chey, W. D.; Lembo, A.; Karsan, H.; Tillisch, K.; Talley, J.; Mayer, E.; Chang, L. Measuring irritable bowel syndrome patient-reported outcomes with an abdominal pain numeric rating scale. Aliment Pharmacol Ther. 2009;30(11-12):1159-1170.

Daily Diary of Gastroparesis Symptoms Questionnaire (GSDD)

McCallum RWL, A.; Esfandyari, T.; Bhandari, B. R.; Ejskjaer, N.; Cosentino, C.; Helton, N.; Mondou, E.; Quinn, J.; Rousseau, F.; T. Z. P. Phase 2b Study Group. Phase 2b, randomized, double-blind 12-week studies of TZP-102, a ghrelin receptor agonist for diabetic gastroparesis. Neurogastroenterol Motil. 2013;25(11):e705-717.

EuroQol-5D Utility Index (EQ-5D)

Walters SJB, John E. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. Qual Life Res. 2005;14(6):1523-1532.

EuroQol-5D visual analogue scale (EQ-5D-VAS)

Coteur GF, B.; Keininger, D. L.; Kosinski, M. Evaluation of the meaningfulness of health-related quality of life improvements as assessed by the SF-36 and the EQ-5D VAS in patients with active Crohn's disease. Aliment Pharmacol Ther. 2009;29(9):1032-1041.

Fecal Incontinence Quality of Life Scale (FIQL) - Dutch version

Bols EMH, H. J.; Berghmans, L. C.; Baeten, C. G.; de Bie, R. A. Responsiveness and interpretability of incontinence severity scores and FIQL in patients with fecal incontinence: a secondary analysis from a randomized controlled trial. Int Urogynecol J Pelvic Floor Dysfunct. 2013;24(3):469-478.

Gastrointestinal Quality of Life Index (GIQLI)

Chan LM, Shamkant; Walker, Rowan; Arns, Wolfgang; Ambuhl, Patrice; Schiavelli, Ruben. Patient-reported gastrointestinal symptom burden and health-related quality of life following conversion from mycophenolate mofetil to enteric-coated mycophenolate sodium. Transplantation. 2006;81(9):1290-1297.

Gastrointestinal Symptom Rating Scale (GSRS)

Chan LM, Shamkant; Walker, Rowan; Arns, Wolfgang; Ambuhl, Patrice; Schiavelli, Ruben. Patient-reported gastrointestinal symptom burden and health-related quality of life following conversion from mycophenolate mofetil to enteric-coated mycophenolate sodium. Transplantation. 2006;81(9):1290-1297.

Gastroparesis Cardinal Symptom Index-Daily Diary (GCSI-DD)

Revicki DAC, M.; Kuo, B.; Szarka, L. A.; McCormack, J.; Parkman, H. P. Evaluating symptom outcomes in gastroparesis clinical trials: validity and responsiveness of the Gastroparesis Cardinal Symptom Index-Daily Diary (GCSI-DD). Neurogastroenterol Motil. 2012;24(5):456-463, e215-456.

Short Form Health Survey 36-Item (SF-36)

Coteur GF, B.; Keininger, D. L.; Kosinski, M. Evaluation of the meaningfulness of health-related quality of life improvements as assessed by the SF-36 and the EQ-5D VAS in patients with active Crohn's disease. Aliment Pharmacol Ther. 2009;29(9):1032-1041.

Short-Form Six-Dimension (SF-6D)

Walters SJB, John E. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. Qual Life Res. 2005;14(6):1523-1532.

Vaizey score

Bols EMJH, E. J. M.; Deutekom, M.; Berghmans, B. C. M.; Baeten, C. G. M. I.; De Bie, R. A. Inconclusive psychometric properties of the vaizey score in fecally incontinent patients: A prospective cohort study. Neurourology and Urodynamics. 2010;29(3):370-377.

Gastroenterology, Rheumatology

University of California Los Angeles Scleroderma Clinical Trial Consortium Gastrointestinal Tract 2.0 (UCLA SCTC 2.0) Instrument

Khanna DF, Daniel E.; Maranian, Paul; Seibold, James R.; Impens, Ann; Mayes, Maureen D.; Clements, Philip J.; Getzug, Terri; Hays, Ron D. Minimally important differences of the UCLA Scleroderma Clinical Trial Consortium Gastrointestinal Tract Instrument. J Rheumatol. 2011;38(9):1920-1924.

General Surgery

EuroQol-5D Utility Index (EQ-5D)

Browne JPvdM, Jan H.; Lewsey, James D.; Lamping, Donna L.; Black, Nick. Mathematical coupling may account for the association between baseline severity and minimally important difference values. J Clin Epidemiol. 2010;63(8):865-874.

Gastrointestinal Quality of Life Index (GIQLI)

Shi H-YL, King-Teh; Lee, Hao-Hsien; Uen, Yih-Huei; Na, Hsueh-Li; Chao, Fang-Tse; Chiu, Chong-Chi. The minimal clinically important difference in the Gastrointestinal Quality-of-Life Index after cholecystectomy. Surg Endosc. 2009;23(12):2708-2712.

Short Form Health Survey 36-Item (SF-36) - UK version

Browne JPvdM, Jan H.; Lewsey, James D.; Lamping, Donna L.; Black, Nick. Mathematical coupling may account for the association between baseline severity and minimally important difference values. J Clin Epidemiol. 2010;63(8):865-874.

Hematology

Mean Symptom Complex Severity (MSCS) score

Vernon MKR, Anne M.; Wyrwich, Kathleen W.; White, Martha V.; Grienenberger, Aurelie. Psychometric validation of two patient-reported outcome measures to assess symptom severity and changes in symptoms in hereditary angioedema. Qual Life Res. 2009;18(7):929-939.

Pain visual analogue scale (VAS)

Lopez BLF, Pamela; Davis-Moon, Linda; Corbin, Theodore; Ballas, Samir K. Clinically significant differences in the visual analog pain scale in acute vasoocclusive sickle cell crisis. Hemoglobin. 2007;31(4):427-432.

Treatment Outcome Score (TOS)

Vernon MKR, Anne M.; Wyrwich, Kathleen W.; White, Martha V.; Grienenberger, Aurelie. Psychometric validation of two patient-reported outcome measures to assess symptom severity and changes in symptoms in hereditary angioedema. Qual Life Res. 2009;18(7):929-939.

Infectious Disease

Assessment of Body Change and Distress (ABCD) questionnaire

Duracinsky ML, P.; Armstrong, A. R.; Dolivo, M.; Mouly, F.; Chassany, O. A longitudinal evaluation of the impact of a polylactic acid injection therapy on health related quality of life amongst HIV patients treated with anti-retroviral agents under real conditions of use. BMC Infect Dis. 2013;13:92.

Hepatitis C virus patient-reported outcomes (HCV-PRO) instrument

Anderson RTB, R. W.; Erickson, P.; Revicki, D. A.; Dietz, B.; Gooch, K. Psychometric evaluation of the hepatitis C virus patient-reported outcomes (HCV-PRO) instrument: validity, responsiveness, and identification of the minimally important difference in a phase 2 clinical trial. Qual Life Res. 2014;23(3):877-886.

Wisconsin Upper Respiratory Symptom Survey (WURSS-21)

Barrett BB, Roger; Mundt, Marlon. Comparison of anchor-based and distributional approaches in estimating important difference in common cold. Qual Life Res. 2008;17(1):75-85.

Barrett BB, Roger L.; Mundt, Marlon P.; Thomas, Gay R.; Barlow, Shari K.; Highstrom, Alex D.; Bahrainian, Mozhdeh. Validation of a short form Wisconsin Upper Respiratory Symptom Survey (WURSS-21). Health Qual Life Outcomes. 2009;7:76.

Wisconsin Upper Respiratory Symptom Survey (WURSS-44)

Barrett BB, Roger; Mundt, Marlon. Comparison of anchor-based and distributional approaches in estimating important difference in common cold. Qual Life Res. 2008;17(1):75-85.

Barrett BB, Roger L.; Mundt, Marlon P.; Thomas, Gay R.; Barlow, Shari K.; Highstrom, Alex D.; Bahrainian, Mozhdeh. Validation of a short form Wisconsin Upper Respiratory Symptom Survey (WURSS-21). Health Qual Life Outcomes. 2009;7:76.

Musculoskeletal and Chronic Pain (Multi-Disciplinary)

Arm Pain

Pain intensity numerical rating scale (NRS)

Kovacs FMA, V.; Royuela, A.; Corcoll, J.; Alegre, L.; Tomas, M.; Mir, M. A.; Cano, A.; Muriel, A.; Zamora, J.; Del Real, M. T. G.; Gestoso, M.; Mufraggi, N. Minimum detectable and minimal clinically important changes for pain in patients with nonspecific neck pain. BMC Musculoskelet Disord. 2008;9(43).

Patient-Specific functional scale (PSFS)

Abbott JHS, John. Minimum Important Differences for the Patient-Specific Functional Scale, 4 Region-Specific Outcome Measures, and the Numeric Pain Rating Scale. J Orthop Sports Phys Ther. 2014;44(8):560-564.

Upper Extremity Functional Index (UEFI)

Abbott JHS, John. Minimum Important Differences for the Patient-Specific Functional Scale, 4 Region-Specific Outcome Measures, and the Numeric Pain Rating Scale. J Orthop Sports Phys Ther. 2014;44(8):560-564.

Back Pain

Bournemouth Questionnaire (BQ)

Hurst HB, Jennifer. Assessing the clinical significance of change scores recorded on subjective outcome measures. J Manipulative Physiol Ther. 2004;27(1):26-35.

Bournemouth Questionnaire (BQ)

Newell DB, Jennifer E. Responsiveness of the Bournemouth questionnaire in determining minimal clinically important change in subgroups of low back pain patients. Spine. 2010;35(19):1801-1806.

Brief Pain Inventory (BPI)

Yeh CHC, L. C.; Chiang, Y. C.; Huang, L. C. Auricular point acupressure for chronic low back pain: A feasibility study for 1-week treatment. Evidence-based Complementary and Alternative Medicine. 2012;2012(383257).

Core Outcome Measures Index (COMI)

Genevay SM, M.; Courvoisier, D. S.; Foltz, V.; Mahieu, G.; Demoulin, C.; Fontana, A. G.; Norberg, M.; de Goumoens, P.; Cedraschi, C.; Rozenberg, S.; Section Rachisde la Societe Francaise de, Rhumatologie. Validity of the French version of the Core Outcome Measures Index for low back pain patients: a prospective cohort study. Eur Spine J. 2014;23(10):2097-2104.

Dartmouth Cooperative (COOP) Charts

Bronfort GB, Lex M. Responsiveness of general health status in chronic low back pain: A comparison of the COOP charts and the SF-36. Pain. 1999;83(2):201-209.

Disease activity numerical rating scale (NRS)

Tubach FR, P.; Martin-Mola, E.; Awada, H.; Bellamy, N.; Bombardier, C.; Felson, D. T.; Hajjaj-Hassouni, N.; Hochberg, M.; Logeart, I.; Matucci-Cerinic, M.; van de Laar, M.; van der Heijde, D.; Dougados, M. Minimum clinically important improvement and patient acceptable symptom state in pain and function in rheumatoid arthritis, ankylosing spondylitis, chronic back pain, hand osteoarthritis, and hip and knee osteoarthritis: Results from a prospective multinational study. Arthritis Care Res (Hoboken). 2012;64(11):1699-1707.

EuroQol-5D Utility Index (EQ-5D)

Walters SJB, John E. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. Qual Life Res. 2005;14(6):1523-1532.

van der Roer NO, Raymond W. J. G.; Bekkering, Geertruida E.; van Tulder, Maurits W.; de Vet, Henrica C. W. Minimal clinically important change for pain intensity, functional status, and general health status in patients with nonspecific low back pain. Spine. 2006;31(5):578-582.

EuroQol-5D Utility Index (EQ-5D) - Dutch weights / Dutch version

Soer RR, M. F.; Speijer, B. L. G. N.; Coppes, M. H.; Vroomen, P. C. A. J. Clinimetric properties of the EuroQol-5D in patients with chronic low back pain. Spine Journal. 2012;12(11):1035-1039.

EuroQoL-5D visual analogue scale (EQ-5D-VAS) - Dutch version

Soer RR, M. F.; Speijer, B. L. G. N.; Coppes, M. H.; Vroomen, P. C. A. J. Clinimetric properties of the EuroQol-5D in patients with chronic low back pain. Spine Journal. 2012;12(11):1035-1039.

Functional disability numerical rating scale (NRS)

Tubach FR, P.; Martin-Mola, E.; Awada, H.; Bellamy, N.; Bombardier, C.; Felson, D. T.; Hajjaj-Hassouni, N.; Hochberg, M.; Logeart, I.; Matucci-Cerinic, M.; van de Laar, M.; van der Heijde, D.; Dougados, M. Minimum clinically important improvement and patient acceptable symptom state in pain and function in rheumatoid arthritis, ankylosing spondylitis, chronic back pain, hand osteoarthritis, and hip and knee osteoarthritis: Results from a prospective multinational study. Arthritis Care Res (Hoboken). 2012;64(11):1699-1707.

Functional rating index (FRI)

Childs JDP, Sara R. Psychometric properties of the functional rating index in patients with low back pain.[Erratum appears in Eur Spine J. 2005 Dec;14(10):1013]. Eur Spine J. 2005;14(10):1008-1012.

Hannover Functional Ability Questionnaire (FFbH-R)

Strand LIA, Bodil; Lygren, Hildegunn; Skouen, Jan Sture; Ostelo, Raymond; Magnussen, Liv Heide. Responsiveness to change of 10 physical tests used for patients with back pain. Phys Ther. 2011;91(3):404-415.

Low-Back Short Form-36 Physical Functioning 18 (SF-36 PF18)

Davidson MK, Jennifer L.; Eyres, Sophie. A low back-specific version of the SF-36 Physical Functioning scale. Spine. 2004;29(5):586-594.

modified Von Korff Scales

Froud RA, G. Using ROC curves to choose minimally important change thresholds when sensitivity and specificity are valued equally: The forgotten lesson of pythagoras. Theoretical considerations and an example application of change in health status. PLoS ONE. 2014;9(12).

Oswestry Disability Index (ODI)

Childs JDP, Sara R. Psychometric properties of the functional rating index in patients with low back pain.[Erratum appears in Eur Spine J. 2005 Dec;14(10):1013]. Eur Spine J. 2005;14(10):1008-1012.

de Vet HCB, L. M.; Bezemer, P. D.; Beurskens, A. J. Reproducibility and responsiveness of evaluative outcome measures. Theoretical considerations illustrated by an empirical example. Int J Technol Assess Health Care. 2001;17(4):479-487.

Abbott JHS, John. Minimum Important Differences for the Patient-Specific Functional Scale, 4 Region-Specific Outcome Measures, and the Numeric Pain Rating Scale. J Orthop Sports Phys Ther. 2014;44(8):560-564.

Oswestry Disability Index (ODI) - version 2

Maughan EFL, Jeremy S. Outcome measures in chronic low back pain. Eur Spine J. 2010;19(9):1484-1494.

Pain Disability Index (PDI)

Soer RR, Michiel F.; Vroomen, Patrick C. A. J.; Stegeman, Patrick; Coppes, Maarten H. Responsiveness and minimal clinically important change of the Pain Disability Index in patients with chronic back pain. Spine. 2012;37(8):711-715.

Pain intensity numerical rating scale (NRS)

van der Roer NO, Raymond W. J. G.; Bekkering, Geertruida E.; van Tulder, Maurits W.; de Vet, Henrica C. W. Minimal clinically important change for pain intensity, functional status, and general health status in patients with nonspecific low back pain. Spine. 2006;31(5):578-582.

de Vet HCWO, Raymond W. J. G.; Terwee, Caroline B.; van der Roer, Nicole; Knol, Dirk L.; Beckerman, Heleen; Boers, Maarten; Bouter, Lex M. Minimally important change determined by a visual method integrating an anchor-based and a distribution-based approach. Qual Life Res. 2007;16(1):131-142.

Childs JDP, Sara R.; Fritz, Julie M. Responsiveness of the numeric pain rating scale in patients with low back pain. Spine. 2005;30(11):1331-1334.

Pain numerical rating scale (NRS)

Tubach FR, P.; Martin-Mola, E.; Awada, H.; Bellamy, N.; Bombardier, C.; Felson, D. T.; Hajjaj-Hassouni, N.; Hochberg, M.; Logeart, I.; Matucci-Cerinic, M.; van de Laar, M.; van der Heijde, D.; Dougados, M. Minimum clinically important improvement and patient acceptable symptom state in pain and function in rheumatoid arthritis, ankylosing spondylitis, chronic back pain, hand osteoarthritis, and hip and knee osteoarthritis: Results from a prospective multinational study. Arthritis Care Res (Hoboken). 2012;64(11):1699-1707.

Maughan EFL, Jeremy S. Outcome measures in chronic low back pain. Eur Spine J. 2010;19(9):1484-1494.

Pain self-efficacy questionnaire (PSEQ)

Maughan EFL, Jeremy S. Outcome measures in chronic low back pain. Eur Spine J. 2010;19(9):1484-1494.

Patient-Specific functional scale (PSFS)

Maughan EFL, Jeremy S. Outcome measures in chronic low back pain. Eur Spine J. 2010;19(9):1484-1494.

Abbott JHS, John. Minimum Important Differences for the Patient-Specific Functional Scale, 4 Region-Specific Outcome Measures, and the Numeric Pain Rating Scale. J Orthop Sports Phys Ther. 2014;44(8):560-564.

Quebec Back Pain Disability Scale (QBPDS)

van der Roer NO, Raymond W. J. G.; Bekkering, Geertruida E.; van Tulder, Maurits W.; de Vet, Henrica C. W. Minimal clinically important change for pain intensity, functional status, and general health status in patients with nonspecific low back pain. Spine. 2006;31(5):578-582.

Quebec Back Pain Disability Scale (QBPDS) - Portuguese version

Vieira ACM, S.; Fernandes, R.; Carnide, F.; Cruz, E. B. Responsiveness and interpretability of the Portuguese version of the Quebec Back Pain Disability Scale in patients with chronic low back pain. Spine. 2014;39(5):E346-352.

Roland-Morris Disability Questionnaire (RMDQ)

Jordan KD, Kate M.; Lewis, Martyn; Croft, Peter. A minimal clinically important difference was derived for the Roland-Morris Disability Questionnaire for low back pain. J Clin Epidemiol. 2006;59(1):45-52.

Maughan EFL, Jeremy S. Outcome measures in chronic low back pain. Eur Spine J. 2010;19(9):1484-1494.

de Vet HCB, L. M.; Bezemer, P. D.; Beurskens, A. J. Reproducibility and responsiveness of evaluative outcome measures. Theoretical considerations illustrated by an empirical example. Int J Technol Assess Health Care. 2001;17(4):479-487.

Riddle DLS, P. W.; Binkley, J. M. Sensitivity to change of the Roland-Morris Back Pain Questionnaire: part 2. Phys Ther. 1998;78(11):1197-1207.

Froud RA, G. Using ROC curves to choose minimally important change thresholds when sensitivity and specificity are valued equally: The forgotten lesson of pythagoras. Theoretical considerations and an example application of change in health status. PLoS ONE. 2014;9(12).

Short Form Health Survey 36-Item (SF-36)

Bronfort GB, Lex M. Responsiveness of general health status in chronic low back pain: A comparison of the COOP charts and the SF-36. Pain. 1999;83(2):201-209.

Short-Form Six-Dimension (SF-6D)

Walters SJB, John E. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. Qual Life Res. 2005;14(6):1523-1532.

Back and/or Leg pain

Bournemouth Questionnaire (BQ)

Lauridsen HHM, Claus; Korsholm, Lars; Grunnet-Nilsson, Niels; Hartvigsen, Jan. What is an acceptable outcome of treatment before it begins? Methodological considerations and implications for patients with chronic low back pain. Eur Spine J. 2009;18(12):1858-1866.

Leg pain intensity numerical rating scale (NRS)

Kovacs FMA, Victor; Royuela, Ana; Corcoll, Josep; Alegre, Luis; Cano, Alejandra; Muriel, Alfonso; Zamora, Javier; del Real, Maria Teresa Gil; Gestoso, Mario; Mufraggi, Nicole. Minimal clinically important change for pain intensity and disability in patients with nonspecific low back pain. Spine. 2007;32(25):2915-2920.

Low back pain intensity numerical rating scale (NRS)

Kovacs FMA, Victor; Royuela, Ana; Corcoll, Josep; Alegre, Luis; Cano, Alejandra; Muriel, Alfonso; Zamora, Javier; del Real, Maria Teresa Gil; Gestoso, Mario; Mufraggi, Nicole. Minimal

clinically important change for pain intensity and disability in patients with nonspecific low back pain. Spine. 2007;32(25):2915-2920.

Oswestry Disability Index (ODI) - version 2.1

Lauridsen HHM, Claus; Korsholm, Lars; Grunnet-Nilsson, Niels; Hartvigsen, Jan. What is an acceptable outcome of treatment before it begins? Methodological considerations and implications for patients with chronic low back pain. Eur Spine J. 2009;18(12):1858-1866.

Pain intensity numerical rating scale (NRS)

Kovacs FMA, Victor; Royuela, Ana; Corcoll, Josep; Alegre, Luis; Cano, Alejandra; Muriel, Alfonso; Zamora, Javier; del Real, Maria Teresa Gil; Gestoso, Mario; Mufraggi, Nicole. Minimal clinically important change for pain intensity and disability in patients with nonspecific low back pain. Spine. 2007;32(25):2915-2920.

Pain numerical rating scale (NRS)

Lauridsen HHM, Claus; Korsholm, Lars; Grunnet-Nilsson, Niels; Hartvigsen, Jan. What is an acceptable outcome of treatment before it begins? Methodological considerations and implications for patients with chronic low back pain. Eur Spine J. 2009;18(12):1858-1866.

Roland-Morris Disability Questionnaire (RMDQ)

Kovacs FMA, Victor; Royuela, Ana; Corcoll, Josep; Alegre, Luis; Cano, Alejandra; Muriel, Alfonso; Zamora, Javier; del Real, Maria Teresa Gil; Gestoso, Mario; Mufraggi, Nicole. Minimal clinically important change for pain intensity and disability in patients with nonspecific low back pain. Spine. 2007;32(25):2915-2920.

Chronic Pain (Non-specific)

PROMIS computerized-adaptive test (CAT) Emotional Distress Domain-Anxiety

Swanholm EM, Wade; Makris, Una; Noe, Carl; Gatchel, Robert. Estimates of minimally important differences (MIDs) for two Patient-Reported Outcomes Measurement Information System (PROMIS) computer-adaptive tests in chronic pain patients. Journal of Applied Biobehavioral Research. 2014;19(4):217-232.

PROMIS computerized-adaptive test (CAT) Emotional Distress Domain-Depression

Swanholm EM, Wade; Makris, Una; Noe, Carl; Gatchel, Robert. Estimates of minimally important differences (MIDs) for two Patient-Reported Outcomes Measurement Information System (PROMIS) computer-adaptive tests in chronic pain patients. Journal of Applied Biobehavioral Research. 2014;19(4):217-232.

Knee Pain

Lower Extremity Functional Scale (LEFS) using computerized adaptive test (CAT)

Wang Y-CH, Dennis L.; Stratford, Paul W.; Mioduski, Jerome E. Baseline dependency of minimal clinically important improvement. Phys Ther. 2011;91(5):675-688.

Western Ontario and McMaster Universities Arthritis Index (WOMAC)

Terwee CBR, Leo D.; Dekker, Joost; Bierma-Zeinstra, Sita M.; Peat, George; Jordan, Kelvin P.; Croft, Peter; de Vet, Henrica C. W. Mind the MIC: large variation among populations and methods. J Clin Epidemiol. 2010;63(5):524-534.

Leg Pain

Lower Extremity Functional Scale (LEFS)

Abbott JHS, John. Minimum Important Differences for the Patient-Specific Functional Scale, 4 Region-Specific Outcome Measures, and the Numeric Pain Rating Scale. J Orthop Sports Phys Ther. 2014;44(8):560-564.

Patient-Specific functional scale (PSFS)

Abbott JHS, John. Minimum Important Differences for the Patient-Specific Functional Scale, 4 Region-Specific Outcome Measures, and the Numeric Pain Rating Scale. J Orthop Sports Phys Ther. 2014;44(8):560-564.

Neck Pain

Bournemouth Questionnaire (BQ)

Hurst HB, Jennifer. Assessing the clinical significance of change scores recorded on subjective outcome measures. J Manipulative Physiol Ther. 2004;27(1):26-35.

Neck Disability Index (NDI)

Pool JJMO, Raymond W. J. G.; Hoving, Jan L.; Bouter, Lex M.; de Vet, Henrica C. W. Minimal clinically important change of the Neck Disability Index and the Numerical Rating Scale for patients with neck pain. Spine. 2007;32(26):3047-3051.

Young BAW, Michael J.; Strunce, Joseph B.; Boyles, Robert E.; Whitman, Julie M.; Childs, John D. Responsiveness of the Neck Disability Index in patients with mechanical neck disorders. Spine J. 2009;9(10):802-808.

Jorritsma WD, P. U.; de Vries, G. E.; Geertzen, J. H.; Reneman, M. F. Detecting relevant changes and responsiveness of Neck Pain and Disability Scale and Neck Disability Index. Eur Spine J. 2012;21(12):2550-2557.

Schuller WO, R. W. J. G.; Janssen, R.; de Vet, H. C. W. The influence of study population and definition of improvement on the smallest detectable change and the minimal important change of the neck disability index. Health and Quality of Life Outcomes. 2014;12(1).

Abbott JHS, John. Minimum Important Differences for the Patient-Specific Functional Scale, 4 Region-Specific Outcome Measures, and the Numeric Pain Rating Scale. J Orthop Sports Phys Ther. 2014;44(8):560-564.

Neck Disability Index (NDI) - Norwegian version

Johansen JBR, C.; Bakke, E.; Mengshoel, A. M.; Andelic, N. Reliability and responsiveness of the Norwegian version of the Neck Disability Index. Scandinavian Journal of Pain. 2014;5(1):28-33.

Neck Pain and Disability Scale (NPAD)

Jorritsma WD, P. U.; de Vries, G. E.; Geertzen, J. H.; Reneman, M. F. Detecting relevant changes and responsiveness of Neck Pain and Disability Scale and Neck Disability Index. Eur Spine J. 2012;21(12):2550-2557.

Pain intensity numerical rating scale (NRS)

Pool JJMO, Raymond W. J. G.; Hoving, Jan L.; Bouter, Lex M.; de Vet, Henrica C. W. Minimal clinically important change of the Neck Disability Index and the Numerical Rating Scale for patients with neck pain. Spine. 2007;32(26):3047-3051.

Kovacs FMA, V.; Royuela, A.; Corcoll, J.; Alegre, L.; Tomas, M.; Mir, M. A.; Cano, A.; Muriel, A.; Zamora, J.; Del Real, M. T. G.; Gestoso, M.; Mufraggi, N. Minimum detectable and minimal clinically important changes for pain in patients with nonspecific neck pain. BMC Musculoskelet Disord. 2008;9(43).

Patient-Specific functional scale (PSFS)

Abbott JHS, John. Minimum Important Differences for the Patient-Specific Functional Scale, 4 Region-Specific Outcome Measures, and the Numeric Pain Rating Scale. J Orthop Sports Phys Ther. 2014;44(8):560-564.

Neurology

ABILHAND

Wang T-nL, Keh-chung; Wu, Ching-yi; Chung, Chia-ying; Pei, Yu-cheng; Teng, Yu-kuei. Validity, responsiveness, and clinically important difference of the ABILHAND questionnaire in patients with stroke. Arch Phys Med Rehabil. 2011;92(7):1086-1091.

Amyotrophic Lateral Sclerosis Assessment Questionnaire (ALSAQ-40)

Jenkinson CP, V.; Jones, G.; Fitzpatrick, R. Interpreting change scores on the Amyotrophic Lateral Sclerosis Assessment Questionnaire (ALSAQ-40). Clin Rehabil. 2003;17(4):380-385.

Back &/or buttock symptoms numerical rating scale (NRS)

Cleland JAW, J. M.; Houser, J. L.; Wainner, R. S.; Childs, J. D. Psychometric properties of selected tests in patients with lumbar spinal stenosis. Spine Journal. 2012;12(10):921-931.

Barthel Index (BI)

Hsieh Y-WW, Chun-Hou; Wu, Shwu-Chong; Chen, Pau-Chung; Sheu, Ching-Fan; Hsieh, Ching-Lin. Establishing the minimal clinically important difference of the Barthel Index in stroke patients. Neurorehabil Neural Repair. 2007;21(3):233-238.

Covi Anxiety Scale

Rejas JP, Antonio; Ruiz, Miguel Angel. Standard error of measurement as a valid alternative to minimally important difference for evaluating the magnitude of changes in patient-reported outcomes measures. J Clin Epidemiol. 2008;61(4):350-356.

Disabilities of the Arm, Shoulder, and Hand (DASH)

Sorensen AAH, D.; Tan, W. H.; Ketchersid, J.; Calfee, R. P. Minimal clinically important differences of 3 patient-rated outcomes instruments. J Hand Surg [Am]. 2013;38(4):641-649.

Disabilities of the Arm, Shoulder, and Hand (QuickDASH)

Sorensen AAH, D.; Tan, W. H.; Ketchersid, J.; Calfee, R. P. Minimal clinically important differences of 3 patient-rated outcomes instruments. J Hand Surg [Am]. 2013;38(4):641-649.

Disability and Impact Profile (DIP)

de Groot VB, H.; Uitdehaag, B. M. J.; de Vet, H. C. W.; Lankhorst, G. J.; Polman, C. H.; Bouter, L. M. The usefulness of evaluative outcome measures in patients with multiple sclerosis. Brain. 2006;129(Pt 10):2648-2659.

Fatigue Impact Scale (FIS)

Rendas-Baum RY, Min; Cattelin, Francoise; Wallenstein, Gene V.; Fisk, John D. A novel approach to estimate the minimally important difference for the Fatigue Impact Scale in multiple sclerosis patients. Qual Life Res. 2010;19(9):1349-1358.

Fatigue Severity Scale (FSS) - local language translations

Robinson D, Jr.; Zhao, N.; Gathany, T.; Kim, L. L.; Cella, D.; Revicki, D. Health perceptions and clinical characteristics of relapsing-remitting multiple sclerosis patients: baseline data from an international clinical trial. Curr Med Res Opin. 2009;25(5):1121-1130.

Functional Status Questionnaire (FSQ)

Paltamaa JS, T.; Leskinen, E.; Wikstrom, J.; Malkia, E. Measuring deterioration in international classification of functioning domains of people with multiple sclerosis who are ambulatory. Phys Ther. 2008;88(2):176-190.

Hamburg Quality Of Life Questionnaire Multiple Sclerosis (HAQUAMS)

Gold SMS, Holger; Stein, Heike; Solf, Katrin; Schulz, Karl-Heinz; Heesen, Christoph. Responsiveness of patient-based and external rating scales in multiple sclerosis: head-to-head comparison in three clinical settings. J Neurol Sci. 2010;290(1-2):102-106.

Headache Impact Test (HIT-6)

Coeytaux RRK, Jay S.; Chao, Ryon; Mann, J. Douglas; Devellis, Robert F. Four methods of estimating the minimal important difference score were compared to establish a clinically significant change in Headache Impact Test. J Clin Epidemiol. 2006;59(4):374-380.

Smelt AFA, W. J.; Terwee, C. B.; Ferrari, M. D.; Blom, J. W. What is a clinically relevant change on the HIT-6 questionnaire? An estimation in a primary-care population of migraine patients. Cephalalgia. 2014;34(1):29-36.

Medical Outcomes Study (MOS) Sleep Scale - global index of sleep interference

Rejas JP, Antonio; Ruiz, Miguel Angel. Standard error of measurement as a valid alternative to minimally important difference for evaluating the magnitude of changes in patient-reported outcomes measures. J Clin Epidemiol. 2008;61(4):350-356.

Modified Oswestry Disability Index (ODI)

Cleland JAW, J. M.; Houser, J. L.; Wainner, R. S.; Childs, J. D. Psychometric properties of selected tests in patients with lumbar spinal stenosis. Spine Journal. 2012;12(10):921-931.

Modified Swiss Spinal Stenosis Scale (SSS)

Cleland JAW, J. M.; Houser, J. L.; Wainner, R. S.; Childs, J. D. Psychometric properties of selected tests in patients with lumbar spinal stenosis. Spine Journal. 2012;12(10):921-931.

Motor Activity Log (MAL)

Lang CEE, Dorothy F.; Birkenmeier, Rebecca L.; Dromerick, Alexander W. Estimating minimal clinically important differences of upper-extremity measures early after stroke. Arch Phys Med Rehabil. 2008;89(9):1693-1700.

Multiple Sclerosis Impact Scale-29 (MSIS)

Schwartz CEA, A.; Motl, R. W. Investigating the minimal important difference in ambulation in multiple sclerosis: A disconnect between performance-based and patient-reported outcomes? J Neurol Sci. 2014;347(1-2):268-274.

Multiple Sclerosis Quality of Life Inventory (MSQLI) - local language translations

Robinson D, Jr.; Zhao, N.; Gathany, T.; Kim, L. L.; Cella, D.; Revicki, D. Health perceptions and clinical characteristics of relapsing-remitting multiple sclerosis patients: baseline data from an international clinical trial. Curr Med Res Opin. 2009;25(5):1121-1130.

Multiple Sclerosis Walking Scale-12 (MSWS)

Schwartz CEA, A.; Motl, R. W. Investigating the minimal important difference in ambulation in multiple sclerosis: A disconnect between performance-based and patient-reported outcomes? J Neurol Sci. 2014;347(1-2):268-274.

Baert IF, J.; Smedal, T.; Dalgas, U.; Romberg, A.; Kalron, A.; Conyers, H.; Elorriaga, I.; Gebara, B.; Gumse, J.; Heric, A.; Jensen, E.; Jones, K.; Knuts, K.; Maertens De Noordhout, B.; Martic, A.; Normann, B.; Eijnde, B. O.; Rasova, K.; Santoyo Medina, C.; Truyens, V.; Wens, I.; Feys, P. Responsiveness and clinically meaningful improvement, according to disability level, of five walking measures after rehabilitation in multiple sclerosis: A European multicenter study. Neurorehabilitation and Neural Repair. 2014;28(7):621-631.

Neurological Fatigue Index for multiple sclerosis (NFI-MS)

Mills RJC, M.; Tennant, A.; Young, C. A. Perceived changes and minimum clinically important difference of the Neurological Fatigue Index for multiple sclerosis (NFI-MS). Mult Scler. 2013;19(4):502-505.

Parkinson's Disease Questionnaire (PDQ-39)

Peto VJ, C.; Fitzpatrick, R. Determining minimally important differences for the PDQ-39 Parkinson's disease questionnaire. Age Ageing. 2001;30(4):299-302.

Parkinson's Disease Questionnaire (PDQ-8) - English or Chinese version

Luo NT, Louis C. S.; Zhao, Yingjiao; Lau, Puay-Ngoh; Au, Wing-Lok; Li, Shu Chuen. Determination of the longitudinal validity and minimally important difference of the 8-item Parkinson's Disease Questionnaire (PDQ-8). Mov Disord. 2009;24(2):183-187.

Patient Assessment of Multiple Sclerosis Impact (PAMSI)

Robinson D, Jr.; Zhao, N.; Gathany, T.; Kim, L. L.; Cella, D.; Revicki, D. Health perceptions and clinical characteristics of relapsing-remitting multiple sclerosis patients: baseline data from an international clinical trial. Curr Med Res Opin. 2009;25(5):1121-1130.

Patient Reported Indices for Multiple Sclerosis (PRIMUS)

Twiss JD, L. C.; McKenna, S. P.; Eckert, B. Interpreting scores on multiple sclerosis-specific patient reported outcome measures (the PRIMUS and U-FIS). Health and Quality of Life Outcomes. 2010;8(117).

Patient reported measures of functional status (FS) using computerized adaptive test (CAT)

Wang Y-CH, Dennis L.; Werneke, Mark; Stratford, Paul W.; Mioduski, Jerome E. Clinical interpretation of outcome measures generated from a lumbar computerized adaptive test. Phys Ther. 2010;90(9):1323-1335.

Patient-Rated Wrist Evaluation (PRWE)

Sorensen AAH, D.; Tan, W. H.; Ketchersid, J.; Calfee, R. P. Minimal clinically important differences of 3 patient-rated outcomes instruments. J Hand Surg [Am]. 2013;38(4):641-649.

Patient-Specific functional scale (PSFS)

Cleland JAW, J. M.; Houser, J. L.; Wainner, R. S.; Childs, J. D. Psychometric properties of selected tests in patients with lumbar spinal stenosis. Spine Journal. 2012;12(10):921-931.

Profile of Mood States (POMS)

Cramer JAH, Anne E.; Kustra, Robert P. Improved mood states with lamotrigine in patients with epilepsy. Epilepsy Behav. 2004;5(5):702-707.

Raskin Depression Scale

Rejas JP, Antonio; Ruiz, Miguel Angel. Standard error of measurement as a valid alternative to minimally important difference for evaluating the magnitude of changes in patient-reported outcomes measures. J Clin Epidemiol. 2008;61(4):350-356.

Rivermead Mobility Index (RMI)

de Groot VB, H.; Uitdehaag, B. M. J.; de Vet, H. C. W.; Lankhorst, G. J.; Polman, C. H.; Bouter, L. M. The usefulness of evaluative outcome measures in patients with multiple sclerosis. Brain. 2006;129(Pt 10):2648-2659.

Rotterdam handicap scale (RHS)

Merkies ISJvN, S. I.; Hanna, K.; Hughes, R. A. C.; Deng, C. Confirming the efficacy of intravenous immunoglobulin in CIDP through minimum clinically important differences: shifting from statistical significance to clinical relevance. J Neurol Neurosurg Psychiatry. 2010;81(11):1194-1199.

SATIS-Stroke questionnaire

Bouffioulx EA, Carlyne; Vandervelde, Laure; Thonnard, Jean-Louis. Changes in satisfaction with activities and participitation between acute, post-acute and chronic stroke phases: a responsiveness study of the SATIS-Stroke questionnaire. J Rehabil Med. 2010;42(10):944-948.

Schwab and England Activities of Daily Living Scale

Martinez-Martin PP, Luis; Forjaz, Maria Joao. Longitudinal metric properties of disability rating scales for Parkinson's disease. Value Health. 2006;9(6):386-393.

Sheehan Disability Scale (SDS)

Rejas JP, Antonio; Ruiz, Miguel Angel. Standard error of measurement as a valid alternative to minimally important difference for evaluating the magnitude of changes in patient-reported outcomes measures. J Clin Epidemiol. 2008;61(4):350-356.

Short Form Health Survey 36-Item (SF-36)

Merkies ISJvN, S. I.; Hanna, K.; Hughes, R. A. C.; Deng, C. Confirming the efficacy of intravenous immunoglobulin in CIDP through minimum clinically important differences: shifting

from statistical significance to clinical relevance. J Neurol Neurosurg Psychiatry. 2010;81(11):1194-1199.

Robinson D, Jr.; Zhao, N.; Gathany, T.; Kim, L. L.; Cella, D.; Revicki, D. Health perceptions and clinical characteristics of relapsing-remitting multiple sclerosis patients: baseline data from an international clinical trial. Curr Med Res Opin. 2009;25(5):1121-1130.

de Groot VB, H.; Uitdehaag, B. M. J.; de Vet, H. C. W.; Lankhorst, G. J.; Polman, C. H.; Bouter, L. M. The usefulness of evaluative outcome measures in patients with multiple sclerosis. Brain. 2006;129(Pt 10):2648-2659.

Short-Form Six-Dimension (SF-6D)

Lee BBK, Madeleine T.; Simpson, Judy M.; Haran, Mark J.; Stockler, Martin R.; Marial, Obaydullah; Salkeld, Glenn. Validity, responsiveness, and minimal important difference for the SF-6D health utility scale in a spinal cord injured population. Value Health. 2008;11(4):680-688.

Spasticity numerical rating scale (NRS)

Farrar JTT, Andrea B.; Stott, Colin; Duncombe, Paul; Jensen, Mark P. Validity, reliability, and clinical importance of change in a 0-10 numeric rating scale measure of spasticity: a post hoc analysis of a randomized, double-blind, placebo-controlled trial. Clin Ther. 2008;30(5):974-985.

Stroke Impact Scale (SIS)

Lin K-cF, Tiffany; Wu, Ching-yi; Wang, Yen-ho; Liu, Jung-sen; Hsieh, Ching-ju; Lin, Shih-fan. Minimal detectable change and clinically important difference of the Stroke Impact Scale in stroke patients. Neurorehabil Neural Repair. 2010;24(5):486-492.

Stroke Impact Scale-16 (SIS-16)

Fulk GDL, Miriam; Dunning, Kari; Golden, Sue; Boyne, Pierce; West, Trent. How much change in the stroke impact scale-16 is important to people who have experienced a stroke? Top. 2010;17(6):477-483.

Thigh/leg symptoms numerical rating scale (NRS)

Cleland JAW, J. M.; Houser, J. L.; Wainner, R. S.; Childs, J. D. Psychometric properties of selected tests in patients with lumbar spinal stenosis. Spine Journal. 2012;12(10):921-931.

Unidimensional Fatigue Impact scale (U-FIS)

Twiss JD, L. C.; McKenna, S. P.; Eckert, B. Interpreting scores on multiple sclerosis-specific patient reported outcome measures (the PRIMUS and U-FIS). Health and Quality of Life Outcomes. 2010;8(117).

Unified Parkinson's Disease Rating Scale (UPDRS)

Schrag AS, Cristina; Counsell, Nicholas; Poewe, Werner. Minimal clinically important change on the unified Parkinson's disease rating scale. Mov Disord. 2006;21(8):1200-1207.

Neurology, Neurosurgery

Activity Impairment Assessment (AIA)

Quadri NL, A.; Keating, K. N.; Nafees, B.; Piccirillo, J.; Wild, D. Psychometric evaluation of the Sinonasal Outcome Test-16 and activity impairment assessment in acute bacterial sinusitis. Otolaryngol Head Neck Surg. 2013;149(1):161-167.

Sino-Nasal Outcome Test-16 (SNOT-16)

Quadri NL, A.; Keating, K. N.; Nafees, B.; Piccirillo, J.; Wild, D. Psychometric evaluation of the Sinonasal Outcome Test-16 and activity impairment assessment in acute bacterial sinusitis. Otolaryngol Head Neck Surg. 2013;149(1):161-167.

Neurosurgery

Barrow Neurological Institute Pain Scale (BNI-PS)

Reddy VKP, S. L.; Lockney, D. T.; Patrawala, S. A.; Su, P. F.; Mericle, R. A. Percutaneous stereotactic radiofrequency lesioning for trigeminal neuralgia: determination of minimum clinically important difference in pain improvement for patient-reported outcomes. Neurosurgery. 2014;74(3):262-266; discussion 266.

Reddy VKP, S. L.; Patrawala, S. A.; Lockney, D. T.; Su, P. F.; Mericle, R. A. Microvascular decompression for classic trigeminal neuralgia: determination of minimum clinically important difference in pain improvement for patient reported outcomes. Neurosurgery. 2013;72(5):749-754; discussion 754.

EuroQol-5D Utility Index (EQ-5D)

Parker SLG, S. S.; Zuckerman, S. L.; Mendenhall, S. K.; Wells, J. A.; Shau, D. N.; McGirt, M. J. Comprehensive assessment of 1-year outcomes and determination of minimum clinically important difference in pain, disability, and quality of life after suboccipital decompression for Chiari malformation I in adults. Neurosurgery. 2013;73(4):569-581; discussion 581.

Head pain numerical rating scale (NRS)

Parker SLG, S. S.; Zuckerman, S. L.; Mendenhall, S. K.; Wells, J. A.; Shau, D. N.; McGirt, M. J. Comprehensive assessment of 1-year outcomes and determination of minimum clinically important difference in pain, disability, and quality of life after suboccipital decompression for Chiari malformation I in adults. Neurosurgery. 2013;73(4):569-581; discussion 581.

Headache Disability Index (HDI)

Parker SLG, S. S.; Zuckerman, S. L.; Mendenhall, S. K.; Wells, J. A.; Shau, D. N.; McGirt, M. J. Comprehensive assessment of 1-year outcomes and determination of minimum clinically important difference in pain, disability, and quality of life after suboccipital decompression for Chiari malformation I in adults. Neurosurgery. 2013;73(4):569-581; discussion 581.

modified Japanese Orthopaedic Association (mJOA)

Parker SLG, S. S.; Zuckerman, S. L.; Mendenhall, S. K.; Wells, J. A.; Shau, D. N.; McGirt, M. J. Comprehensive assessment of 1-year outcomes and determination of minimum clinically important difference in pain, disability, and quality of life after suboccipital decompression for Chiari malformation I in adults. Neurosurgery. 2013;73(4):569-581; discussion 581.

Neck Disability Index (NDI)

Parker SLG, S. S.; Zuckerman, S. L.; Mendenhall, S. K.; Wells, J. A.; Shau, D. N.; McGirt, M. J. Comprehensive assessment of 1-year outcomes and determination of minimum clinically important difference in pain, disability, and quality of life after suboccipital decompression for Chiari malformation I in adults. Neurosurgery. 2013;73(4):569-581; discussion 581.

Neck pain numerical rating scale (NRS)

Parker SLG, S. S.; Zuckerman, S. L.; Mendenhall, S. K.; Wells, J. A.; Shau, D. N.; McGirt, M. J. Comprehensive assessment of 1-year outcomes and determination of minimum clinically important difference in pain, disability, and quality of life after suboccipital decompression for Chiari malformation I in adults. Neurosurgery. 2013;73(4):569-581; discussion 581.

Pain visual analogue scale (VAS)

Reddy VKP, S. L.; Lockney, D. T.; Patrawala, S. A.; Su, P. F.; Mericle, R. A. Percutaneous stereotactic radiofrequency lesioning for trigeminal neuralgia: determination of minimum clinically important difference in pain improvement for patient-reported outcomes. Neurosurgery. 2014;74(3):262-266; discussion 266.

Reddy VKP, S. L.; Patrawala, S. A.; Lockney, D. T.; Su, P. F.; Mericle, R. A. Microvascular decompression for classic trigeminal neuralgia: determination of minimum clinically important difference in pain improvement for patient reported outcomes. Neurosurgery. 2013;72(5):749-754; discussion 754.

Short Form Health Survey 12-Item (SF-12)

Parker SLG, S. S.; Zuckerman, S. L.; Mendenhall, S. K.; Wells, J. A.; Shau, D. N.; McGirt, M. J. Comprehensive assessment of 1-year outcomes and determination of minimum clinically important difference in pain, disability, and quality of life after suboccipital decompression for Chiari malformation I in adults. Neurosurgery. 2013;73(4):569-581; discussion 581.

Zung Self-rating Depression Scale (ZDS)

Parker SLG, S. S.; Zuckerman, S. L.; Mendenhall, S. K.; Wells, J. A.; Shau, D. N.; McGirt, M. J. Comprehensive assessment of 1-year outcomes and determination of minimum clinically important difference in pain, disability, and quality of life after suboccipital decompression for Chiari malformation I in adults. Neurosurgery. 2013;73(4):569-581; discussion 581.

Neurosurgery, Orthopedic Surgery

Arm pain numerical rating scale (NRS)

Carreon LYG, Steven D.; Campbell, Mitchell J.; Anderson, Paul A. Neck Disability Index, short form-36 physical component summary, and pain scales for neck and arm pain: the minimum clinically important difference and substantial clinical benefit after cervical spine fusion. Spine J. 2010;10(6):469-474.

Back pain numerical rating scale (NRS)

Copay AGG, Steven D.; Subach, Brian R.; Berven, Sigurd; Schuler, Thomas C.; Carreon, Leah Y. Minimum clinically important difference in lumbar spine surgery patients: a choice of methods using the Oswestry Disability Index, Medical Outcomes Study questionnaire Short Form 36, and pain scales. Spine J. 2008;8(6):968-974.

Back pain visual analogue scale (VAS)

Hagg OF, P.; Nordwall, A.; Swedish Lumbar Spine Study, Group. The clinical importance of changes in outcome scores after treatment for chronic low back pain. Eur Spine J. 2003;12(1):12-20.

Parker SLM, Stephen K.; Shau, David N.; Adogwa, Owoicho; Anderson, William N.; Devin, Clinton J.; McGirt, Matthew J. Minimum clinically important difference in pain, disability, and quality of life after neural decompression and fusion for same-level recurrent lumbar stenosis: understanding clinical versus statistical significance. J Neurosurg Spine. 2012;16(5):471-478.

Parker SLA, Owoicho; Paul, Alexandra R.; Anderson, William N.; Aaronson, Oran; Cheng, Joseph S.; McGirt, Matthew J. Utility of minimum clinically important difference in assessing pain, disability, and health state after transforaminal lumbar interbody fusion for degenerative lumbar spondylolisthesis. J Neurosurg Spine. 2011;14(5):598-604.

Parker SLA, O.; Mendenhall, S. K.; Shau, D. N.; Anderson, W. N.; Cheng, J. S.; Devin, C. J.; McGirt, M. J. Determination of minimum clinically important difference (MCID) in pain, disability, and quality of life after revision fusion for symptomatic pseudoarthrosis. Spine Journal. 2012;12(12):1122-1128.

Cervical Spine Outcomes Questionnaire (CSOQ)

Skolasky RLA, Todd J.; Maggard, Anica M.; Riley, Lee H., 3rd. Minimum clinically important differences in the Cervical Spine Outcomes Questionnaire: results from a national multicenter study of patients treated with anterior cervical decompression and arthrodesis. J Bone Joint Surg Am. 2011;93(14):1294-1300.

Core Outcome Measures Index (COMI)

Mannion AFP, F.; Kleinstuck, F. S.; Lattig, F.; Jeszenszky, D.; Bartanusz, V.; Dvorak, J.; Grob, D. The quality of spine surgery from the patient's perspective: part 2. Minimal clinically important difference for improvement and deterioration as measured with the Core Outcome Measures Index. Eur Spine J. 2009;18 Suppl 3:374-379.

EuroQol-5D Utility Index (EQ-5D)

Parker SLM, Stephen K.; Shau, David N.; Adogwa, Owoicho; Anderson, William N.; Devin, Clinton J.; McGirt, Matthew J. Minimum clinically important difference in pain, disability, and quality of life after neural decompression and fusion for same-level recurrent lumbar stenosis: understanding clinical versus statistical significance. J Neurosurg Spine. 2012;16(5):471-478.

Parker SLA, Owoicho; Paul, Alexandra R.; Anderson, William N.; Aaronson, Oran; Cheng, Joseph S.; McGirt, Matthew J. Utility of minimum clinically important difference in assessing pain, disability, and health state after transforaminal lumbar interbody fusion for degenerative lumbar spondylolisthesis. J Neurosurg Spine. 2011;14(5):598-604.

Parker SLA, O.; Mendenhall, S. K.; Shau, D. N.; Anderson, W. N.; Cheng, J. S.; Devin, C. J.; McGirt, M. J. Determination of minimum clinically important difference (MCID) in pain, disability, and quality of life after revision fusion for symptomatic pseudoarthrosis. Spine Journal. 2012;12(12):1122-1128.

EuroQol-5D Utility Index (EQ-5D) - US weights

Parker SLM, Stephen K.; Shau, David; Adogwa, Owoicho; Cheng, Joseph S.; Anderson, William N.; Devin, Clinton J.; McGirt, Matthew J. Determination of minimum clinically important difference in pain, disability, and quality of life after extension of fusion for adjacent-segment disease. J Neurosurg Spine. 2012;16(1):61-67.

General Function Score (GFS)

Hagg OF, P.; Nordwall, A.; Swedish Lumbar Spine Study, Group. The clinical importance of changes in outcome scores after treatment for chronic low back pain. Eur Spine J. 2003;12(1):12-20.

Leg pain numerical rating scale (NRS)

Copay AGG, Steven D.; Subach, Brian R.; Berven, Sigurd; Schuler, Thomas C.; Carreon, Leah Y. Minimum clinically important difference in lumbar spine surgery patients: a choice of methods using the Oswestry Disability Index, Medical Outcomes Study questionnaire Short Form 36, and pain scales. Spine J. 2008;8(6):968-974.

Leg pain visual analogue scale (VAS)

Parker SLM, Stephen K.; Shau, David; Adogwa, Owoicho; Cheng, Joseph S.; Anderson, William N.; Devin, Clinton J.; McGirt, Matthew J. Determination of minimum clinically important difference in pain, disability, and quality of life after extension of fusion for adjacent-segment disease. J Neurosurg Spine. 2012;16(1):61-67.

Parker SLM, Stephen K.; Shau, David N.; Adogwa, Owoicho; Anderson, William N.; Devin, Clinton J.; McGirt, Matthew J. Minimum clinically important difference in pain, disability, and quality of life after neural decompression and fusion for same-level recurrent lumbar stenosis: understanding clinical versus statistical significance. J Neurosurg Spine. 2012;16(5):471-478.

Parker SLA, Owoicho; Paul, Alexandra R.; Anderson, William N.; Aaronson, Oran; Cheng, Joseph S.; McGirt, Matthew J. Utility of minimum clinically important difference in assessing pain, disability, and health state after transforaminal lumbar interbody fusion for degenerative lumbar spondylolisthesis. J Neurosurg Spine. 2011;14(5):598-604.

Patel MSN, M.; Sell, P. A comparison of patient-reported outcome measures after spinal surgery. Bone & Joint Journal (British). 2015;97-B(3):366-371.

Low-back pain visual analogue scale (VAS)

Parker SLM, Stephen K.; Shau, David; Adogwa, Owoicho; Cheng, Joseph S.; Anderson, William N.; Devin, Clinton J.; McGirt, Matthew J. Determination of minimum clinically important difference in pain, disability, and quality of life after extension of fusion for adjacent-segment disease. J Neurosurg Spine. 2012;16(1):61-67.

Maine Seattle Back Questionnaire (MSBQ)

Grovle LH, A. J.; Hasvik, E.; Natvig, B.; Brox, J. I.; Grotle, M. Patients' ratings of global perceived change during 2 years were strongly influenced by the current health status. J Clin Epidemiol. 2014;67(5):508-515.

Neck Disability Index (NDI)

Skolasky RLA, Todd J.; Maggard, Anica M.; Riley, Lee H., 3rd. Minimum clinically important differences in the Cervical Spine Outcomes Questionnaire: results from a national multicenter study of patients treated with anterior cervical decompression and arthrodesis. J Bone Joint Surg Am. 2011;93(14):1294-1300.

Carreon LYG, Steven D.; Campbell, Mitchell J.; Anderson, Paul A. Neck Disability Index, short form-36 physical component summary, and pain scales for neck and arm pain: the minimum clinically important difference and substantial clinical benefit after cervical spine fusion. Spine J. 2010;10(6):469-474.

Auffinger BML, R. R.; Dahdaleh, N. S.; Wong, A. P.; Lam, S. K.; Koski, T.; Fessler, R. G.; Smith, Z. A. Measuring surgical outcomes in cervical spondylotic myelopathy patients undergoing anterior cervical discectomy and fusion: assessment of minimum clinically important difference. PLoS ONE. 2013;8(6):e67408.

Auffinger BL, S.; Shen, J.; Roitberg, B. Z. Measuring surgical outcomes in subaxial degenerative cervical spine disease patients: minimum clinically important difference as a tool for determining meaningful clinical improvement. Neurosurgery. 2014;74(2):206-213; discussion 213-204.

Auffinger BL, S.; Shen, J.; Thaci, B.; Roitberg, B. Z. Usefulness of minimum clinically important difference for assessing patients with subaxial degenerative cervical spine disease: statistical versus substantial clinical benefit. Acta Neurochir (Wien). 2013;155(12):2345-2354; discussion 2355.

Neck pain numerical rating scale (NRS)

Carreon LYG, Steven D.; Campbell, Mitchell J.; Anderson, Paul A. Neck Disability Index, short form-36 physical component summary, and pain scales for neck and arm pain: the minimum clinically important difference and substantial clinical benefit after cervical spine fusion. Spine J. 2010;10(6):469-474.

Neck pain visual analogue scale (VAS)

Auffinger BML, R. R.; Dahdaleh, N. S.; Wong, A. P.; Lam, S. K.; Koski, T.; Fessler, R. G.; Smith, Z. A. Measuring surgical outcomes in cervical spondylotic myelopathy patients undergoing anterior cervical discectomy and fusion: assessment of minimum clinically important difference. PLoS ONE. 2013;8(6):e67408.

Auffinger BL, S.; Shen, J.; Roitberg, B. Z. Measuring surgical outcomes in subaxial degenerative cervical spine disease patients: minimum clinically important difference as a tool for determining meaningful clinical improvement. Neurosurgery. 2014;74(2):206-213; discussion 213-204.

Oswestry Disability Index (ODI)

Hagg OF, P.; Nordwall, A.; Swedish Lumbar Spine Study, Group. The clinical importance of changes in outcome scores after treatment for chronic low back pain. Eur Spine J. 2003;12(1):12-20.

Parker SLM, Stephen K.; Shau, David; Adogwa, Owoicho; Cheng, Joseph S.; Anderson, William N.; Devin, Clinton J.; McGirt, Matthew J. Determination of minimum clinically important difference in pain, disability, and quality of life after extension of fusion for adjacent-segment disease. J Neurosurg Spine. 2012;16(1):61-67.

Parker SLM, Stephen K.; Shau, David N.; Adogwa, Owoicho; Anderson, William N.; Devin, Clinton J.; McGirt, Matthew J. Minimum clinically important difference in pain, disability, and quality of life after neural decompression and fusion for same-level recurrent lumbar stenosis: understanding clinical versus statistical significance. J Neurosurg Spine. 2012;16(5):471-478.

Parker SLA, Owoicho; Paul, Alexandra R.; Anderson, William N.; Aaronson, Oran; Cheng, Joseph S.; McGirt, Matthew J. Utility of minimum clinically important difference in assessing pain, disability, and health state after transforaminal lumbar interbody fusion for degenerative lumbar spondylolisthesis. J Neurosurg Spine. 2011;14(5):598-604.

Parker SLA, O.; Mendenhall, S. K.; Shau, D. N.; Anderson, W. N.; Cheng, J. S.; Devin, C. J.; McGirt, M. J. Determination of minimum clinically important difference (MCID) in pain, disability, and quality of life after revision fusion for symptomatic pseudoarthrosis. Spine Journal. 2012;12(12):1122-1128.

Patel MSN, M.; Sell, P. A comparison of patient-reported outcome measures after spinal surgery. Bone & Joint Journal (British). 2015;97-B(3):366-371.

Oswestry Disability Index (ODI) - version 1

Copay AGG, Steven D.; Subach, Brian R.; Berven, Sigurd; Schuler, Thomas C.; Carreon, Leah Y. Minimum clinically important difference in lumbar spine surgery patients: a choice of methods using the Oswestry Disability Index, Medical Outcomes Study questionnaire Short Form 36, and pain scales. Spine J. 2008;8(6):968-974.

Pain visual analogue scale (VAS)

Grovle LH, A. J.; Hasvik, E.; Natvig, B.; Brox, J. I.; Grotle, M. Patients' ratings of global perceived change during 2 years were strongly influenced by the current health status. J Clin Epidemiol. 2014;67(5):508-515.

Sciatica Bothersomeness Index (SBI)

Grovle LH, A. J.; Hasvik, E.; Natvig, B.; Brox, J. I.; Grotle, M. Patients' ratings of global perceived change during 2 years were strongly influenced by the current health status. J Clin Epidemiol. 2014;67(5):508-515.

Scoliosis Research Society 22 (SRS 22) Patient Questionnaire

Bago JP-G, Francisco J. S.; Les, Esther; Hernandez, Pablo; Pellise, Ferran. Minimal important differences of the SRS-22 Patient Questionnaire following surgical treatment of idiopathic scoliosis. Eur Spine J. 2009;18(12):1898-1904.

Carreon LYS, James O.; Diab, Mohammad; Sucato, Daniel J.; Sturm, Peter F.; Glassman, Steven D.; Spinal Deformity Study, Group. The minimum clinically important difference in Scoliosis Research Society-22 Appearance, Activity, And Pain domains after surgical correction of adolescent idiopathic scoliosis. Spine. 2010;35(23):2079-2083.

Scoliosis Research Society 22R (SRS 22R) Patient Questionnaire

Crawford 3rd CHG, Steven D.; Bridwell, Keith H.; Berven, Sigurd H.; Carreon, Leah Y. The Minimum Clinically Important Difference in SRS-22R Total Score, Appearance, Activity and Pain Domains After Surgical Treatment of Adult Spinal Deformity. Spine. 2015;40(6):377-381.

Short Form Health Survey 12-Item (SF-12)

Parker SLM, Stephen K.; Shau, David; Adogwa, Owoicho; Cheng, Joseph S.; Anderson, William N.; Devin, Clinton J.; McGirt, Matthew J. Determination of minimum clinically important difference in pain, disability, and quality of life after extension of fusion for adjacent-segment disease. J Neurosurg Spine. 2012;16(1):61-67.

Parker SLM, Stephen K.; Shau, David N.; Adogwa, Owoicho; Anderson, William N.; Devin, Clinton J.; McGirt, Matthew J. Minimum clinically important difference in pain, disability, and quality of life after neural decompression and fusion for same-level recurrent lumbar stenosis: understanding clinical versus statistical significance. J Neurosurg Spine. 2012;16(5):471-478.

Parker SLA, O.; Mendenhall, S. K.; Shau, D. N.; Anderson, W. N.; Cheng, J. S.; Devin, C. J.; McGirt, M. J. Determination of minimum clinically important difference (MCID) in pain, disability, and quality of life after revision fusion for symptomatic pseudoarthrosis. Spine Journal. 2012;12(12):1122-1128.

Short Form Health Survey 36-Item (SF-36)

Skolasky RLA, Todd J.; Maggard, Anica M.; Riley, Lee H., 3rd. Minimum clinically important differences in the Cervical Spine Outcomes Questionnaire: results from a national multicenter study of patients treated with anterior cervical decompression and arthrodesis. J Bone Joint Surg Am. 2011;93(14):1294-1300.

Carreon LYG, Steven D.; Campbell, Mitchell J.; Anderson, Paul A. Neck Disability Index, short form-36 physical component summary, and pain scales for neck and arm pain: the minimum clinically important difference and substantial clinical benefit after cervical spine fusion. Spine J. 2010;10(6):469-474.

Auffinger BML, R. R.; Dahdaleh, N. S.; Wong, A. P.; Lam, S. K.; Koski, T.; Fessler, R. G.; Smith, Z. A. Measuring surgical outcomes in cervical spondylotic myelopathy patients undergoing anterior cervical discectomy and fusion: assessment of minimum clinically important difference. PLoS ONE. 2013;8(6):e67408.

Auffinger BL, S.; Shen, J.; Roitberg, B. Z. Measuring surgical outcomes in subaxial degenerative cervical spine disease patients: minimum clinically important difference as a tool for determining meaningful clinical improvement. Neurosurgery. 2014;74(2):206-213; discussion 213-204.

Grovle LH, A. J.; Hasvik, E.; Natvig, B.; Brox, J. I.; Grotle, M. Patients' ratings of global perceived change during 2 years were strongly influenced by the current health status. J Clin Epidemiol. 2014;67(5):508-515.

Auffinger BL, S.; Shen, J.; Thaci, B.; Roitberg, B. Z. Usefulness of minimum clinically important difference for assessing patients with subaxial degenerative cervical spine disease: statistical versus substantial clinical benefit. Acta Neurochir (Wien). 2013;155(12):2345-2354; discussion 2355.

Short Form Health Survey 36-Item (SF-36) - version 2

Copay AGG, Steven D.; Subach, Brian R.; Berven, Sigurd; Schuler, Thomas C.; Carreon, Leah Y. Minimum clinically important difference in lumbar spine surgery patients: a choice of methods using the Oswestry Disability Index, Medical Outcomes Study questionnaire Short Form 36, and pain scales. Spine J. 2008;8(6):968-974.

Zung Self-rating Depression Scale (ZDS)

Hagg OF, P.; Nordwall, A.; Swedish Lumbar Spine Study, Group. The clinical importance of changes in outcome scores after treatment for chronic low back pain. Eur Spine J. 2003;12(1):12-20.

Parker SLM, Stephen K.; Shau, David N.; Adogwa, Owoicho; Anderson, William N.; Devin, Clinton J.; McGirt, Matthew J. Minimum clinically important difference in pain, disability, and quality of life after neural decompression and fusion for same-level recurrent lumbar stenosis: understanding clinical versus statistical significance. J Neurosurg Spine. 2012;16(5):471-478.

Obstetrics and Gynecology

electronic Personal Assessment Questionnaire - Pelvic Floor (ePAQ - PF)

Jones GLR, S. C.; Lumb, J.; Farkas, A. Responsiveness of the electronic Personal Assessment Questionnaire-Pelvic Floor (ePAQ-PF). International Urogynecology Journal and Pelvic Floor Dysfunction. 2009;20(5):557-564.

Endometriosis Health Profile-30 (EHP-30)

Jones GJ, Crispin; Kennedy, Stephen. Evaluating the responsiveness of the Endometriosis Health Profile Questionnaire: the EHP-30. Qual Life Res. 2004;13(3):705-713.

Endometriosis Health Profile-30 (EHP-30) - Dutch version

van de Burgt TJK, K. B.; Hendriks, J. C. Responsiveness of the Dutch Endometriosis Health Profile-30 (EHP-30) questionnaire. Eur J Obstet Gynecol Reprod Biol. 2013;168(1):92-94.

Menorrhagia Impact Questionnaire (MIQ)

Bushnell DMM, Mona L.; Moore, Keith A.; Richter, Holly E.; Rubin, Arkady; Patrick, Donald L. Menorrhagia Impact Questionnaire: assessing the influence of heavy menstrual bleeding on quality of life. Curr Med Res Opin. 2010;26(12):2745-2755.

Pelvic Floor Distress Inventory-20 (PFDI-20)

Utomo EB, B. F.; Steensma, A. B.; Korfage, I. J. Validation of the Pelvic Floor Distress Inventory (PFDI-20) and Pelvic Floor Impact Questionnaire (PFIQ-7) in a Dutch population. Int Urogynecol J Pelvic Floor Dysfunct. 2014;25(4):531-544.

Pelvic Floor Impact Questionnaire-7 (PFIQ-7)

Utomo EB, B. F.; Steensma, A. B.; Korfage, I. J. Validation of the Pelvic Floor Distress Inventory (PFDI-20) and Pelvic Floor Impact Questionnaire (PFIQ-7) in a Dutch population. Int Urogynecol J Pelvic Floor Dysfunct. 2014;25(4):531-544.

Short Form Health Survey 36-Item (SF-36)

Jones GJ, Crispin; Kennedy, Stephen. Evaluating the responsiveness of the Endometriosis Health Profile Questionnaire: the EHP-30. Qual Life Res. 2004;13(3):705-713.

Obstetrics and Gynecology, Urology

Incontinence Quality of Life Instrument (I-QOL)

Patrick DLM, M. L.; Bushnell, D. M.; Yalcin, I.; Wagner, T. H.; Buesching, D. P. Quality of life of women with urinary incontinence: further development of the incontinence quality of life instrument (I-QOL).[Erratum appears in Urology 1999 May;53(5):1072]. Urology. 1999;53(1):71-76.

Overactive Bladder Symptom And Health-Related Quality Of Life Questionnaire (OAB-Q)

Dyer KYX, Yan; Brubaker, Linda; Nygaard, Ingrid; Markland, Alayne; Rahn, David; Chai, Toby C.; Stoddard, Ann; Lukacz, Emily; Urinary Incontinence Treatment, Network. Minimum important difference for validated instruments in women with urge incontinence. Neurourol Urodyn. 2011;30(7):1319-1324.

Pelvic Floor Distress Inventory (PFDI-46)

Barber MDS, Cathie; Janz, Nancy K.; Brubaker, Linda; Nygaard, Ingrid; Nager, Charles W.; Wheeler, Thomas L.; Pelvic Floor Disorders, Network. The minimum important differences for the urinary scales of the Pelvic Floor Distress Inventory and Pelvic Floor Impact Questionnaire. Am J Obstet Gynecol. 2009;200(5):580.e581-587.

Pelvic Floor Distress Inventory (PFDI-46) - Chinese version

Chan SSCC, R. Y. K.; Lai, B. P. Y.; Lee, L. L.; Choy, K. W.; Chung, T. K. H. Responsiveness of the Pelvic Floor Distress Inventory and Pelvic Floor Impact Questionnaire in women undergoing treatment for pelvic floor disorders. International Urogynecology Journal and Pelvic Floor Dysfunction. 2013;24(2):213-221.

Pelvic Floor Impact Questionnaire (PFIQ-93)

Barber MDS, Cathie; Janz, Nancy K.; Brubaker, Linda; Nygaard, Ingrid; Nager, Charles W.; Wheeler, Thomas L.; Pelvic Floor Disorders, Network. The minimum important differences for the urinary scales of the Pelvic Floor Distress Inventory and Pelvic Floor Impact Questionnaire. Am J Obstet Gynecol. 2009;200(5):580.e581-587.

Pelvic Floor Impact Questionnaire (PFIQ-93) - Chinese version

Chan SSCC, R. Y. K.; Lai, B. P. Y.; Lee, L. L.; Choy, K. W.; Chung, T. K. H. Responsiveness of the Pelvic Floor Distress Inventory and Pelvic Floor Impact Questionnaire in women undergoing treatment for pelvic floor disorders. International Urogynecology Journal and Pelvic Floor Dysfunction. 2013;24(2):213-221.

Protection (the use of pads), Amount of urine loss, Frequency of UI, Adjustment (of daily activities or participation due to the UI symptoms), and Body or self-image related to the incontinence symptoms questionnaire (PRAFAB-Q) - Dutch version

Hendriks EJMB, Arnold T. M.; de Bie, Rob A.; de Vet, Henrica C. W. The minimal important change of the PRAFAB questionnaire in women with stress urinary incontinence: results from a prospective cohort study. Neurourol Urodyn. 2008;27(5):379-387.

Urinary Tract Infection Symptom Assessment (UTISA)

Clayson DW, Diane; Doll, Helen; Keating, Karen; Gondek, Kathleen. Validation of a patient-administered questionnaire to measure the severity and bothersomeness of lower urinary tract symptoms in uncomplicated urinary tract infection (UTI): the UTI Symptom Assessment questionnaire. BJU Int. 2005;96(3):350-359.

Urogenital Distress Inventory (UDI)

Dyer KYX, Yan; Brubaker, Linda; Nygaard, Ingrid; Markland, Alayne; Rahn, David; Chai, Toby C.; Stoddard, Ann; Lukacz, Emily; Urinary Incontinence Treatment, Network. Minimum important difference for validated instruments in women with urge incontinence. Neurourol Urodyn. 2011;30(7):1319-1324.

Oncology

15D

Kvam AKF, Peter M.; Wisloff, Finn. Responsiveness and minimal important score differences in quality-of-life questionnaires: a comparison of the EORTC QLQ-C30 cancer-specific questionnaire to the generic utility questionnaires EQ-5D and 15D in patients with multiple myeloma. Eur J Haematol. 2011;87(4):330-337.

8-item index of patient-reported symptoms of renal cell carcinoma (based on Functional Assessment of Cancer Therapy–Biological Response Modifier (FACT-BRM) scale)

Eton DTC, David; Bacik, Jennifer; Motzer, Robert J. A brief symptom index for advanced renal cell carcinoma. Health Qual Life Outcomes. 2006;4:68.

Brief Pain Inventory-Short Form (BPI-SF)

Mathias SDC, Ross D.; Qian, Yi; Jiang, Qi; Dansey, Roger; Chung, Karen. Estimating minimally important differences for the worst pain rating of the Brief Pain Inventory-Short Form. J Support Oncol. 2011;9(2):72-78.

Wong KZ, L.; Zhang, L.; Bedard, G.; Wong, E.; Tsao, M.; Barnes, E.; Danjoux, C.; Sahgal, A.; Holden, L.; Lauzon, N.; Chow, E. Minimal clinically important differences in the brief pain inventory in patients with bone metastases. Support Care Cancer. 2013;21(7):1893-1899.

Cancer Linear Analogue Scale (CLAS)

Patrick DLG, D. D.; Zagari, M. J.; Mathijs, R.; Sweetenham, J.; Epoetin Alfa Study, Group. Assessing the clinical significance of health-related quality of life (HrQOL) improvements in anaemic cancer patients receiving epoetin alfa. Eur J Cancer. 2003;39(3):335-345.

Daily Active Time Exchange (DATE)

Ringash JOS, Brian; Bezjak, Andrea; Redelmeier, Donald A. Interpreting clinically significant changes in patient-reported outcomes. Cancer. 2007;110(1):196-202.

Edmonton Symptom Assessment System (ESAS)

Bedard GZ, L.; Zhang, L.; Lauzon, N.; Holden, L.; Tsao, M.; Danjoux, C.; Barnes, E.; Sahgal, A.; Poon, M.; Chow, E. Minimal clinically important differences in the Edmonton symptom assessment system in patients with advanced cancer. J Pain Symptom Manage. 2013;46(2):192-200.

EORTC Quality of Life Questionnaire - Bone Metastases Module (EORTC QLQ-BM22)

Zeng LC, E.; Zhang, L.; Tseng, L. M.; Hou, M. F.; Fairchild, A.; Vassiliou, V.; Jesus-Garcia, R.; El-Din, M. A. A.; Kumar, A.; Forges, F.; Chie, W. C.; Bedard, G.; Bottomley, A. An international prospective study establishing minimal clinically important differences in the

EORTC QLQ-BM22 and QLQ-C30 in cancer patients with bone metastases. Support Care Cancer. 2012;20(12):3307-3313.

EORTC Quality of Life Questionnaire - Core Questionnaire (EORTC QLQ-C30)

Zeng LC, E.; Zhang, L.; Tseng, L. M.; Hou, M. F.; Fairchild, A.; Vassiliou, V.; Jesus-Garcia, R.; El-Din, M. A. A.; Kumar, A.; Forges, F.; Chie, W. C.; Bedard, G.; Bottomley, A. An international prospective study establishing minimal clinically important differences in the EORTC QLQ-BM22 and QLQ-C30 in cancer patients with bone metastases. Support Care Cancer. 2012;20(12):3307-3313.

Bedard GZ, L.; Zhang, L.; Lauzon, N.; Holden, L.; Tsao, M.; Danjoux, C.; Barnes, E.; Sahgal, A.; Poon, M.; Chow, E. Minimal important differences in the EORTC QLQ-C30 in patients with advanced cancer. Asia-Pacific Journal of Clinical Oncology. 2014;10(2):109-117.

Hong FB, J. L. F.; Bush, N.; Berry, D. L. Patient self-appraisal of change and minimal clinically important difference on the European organization for the research and treatment of cancer quality of life questionnaire core 30 before and during cancer therapy. BMC Cancer. 2013;13(165).

Kvam AKW, Finn; Fayers, Peter M. Minimal important differences and response shift in health-related quality of life; a longitudinal study in patients with multiple myeloma. Health Qual Life Outcomes. 2010;8:79.

Kvam AKF, Peter M.; Wisloff, Finn. Responsiveness and minimal important score differences in quality-of-life questionnaires: a comparison of the EORTC QLQ-C30 cancer-specific questionnaire to the generic utility questionnaires EQ-5D and 15D in patients with multiple myeloma. Eur J Haematol. 2011;87(4):330-337.

Kvam AKF, Peter; Wisloff, Finn. What changes in health-related quality of life matter to multiple myeloma patients? A prospective study. Eur J Haematol. 2010;84(4):345-353.

EORTC Quality of Life Questionnaire - Core Questionnaire (EORTC QLQ-C30) - version 3

Maringwa J, Quinten C, King M, et al. Minimal clinically meaningful differences for the EORTC QLQ-C30 and EORTC QLQ-BN20 scales in brain cancer patients. Ann Oncol. 2011;22(9):2107-2112.

EORTC Quality of Life Questionnaire - Core Questionnaire (EORTC QLQ-C30); Question 29

Doyle CC, M.; Pintilie, M.; Oza, A. M. Does palliative chemotherapy palliate? Evaluation of expectations, outcomes, and costs in women receiving chemotherapy for advanced ovarian cancer. J Clin Oncol. 2001;19(5):1266-1274.

EORTC Quality of Life Questionnaire - Core Questionnaire (EORTC QLQ-C30); Question 30

Doyle CC, M.; Pintilie, M.; Oza, A. M. Does palliative chemotherapy palliate? Evaluation of expectations, outcomes, and costs in women receiving chemotherapy for advanced ovarian cancer. J Clin Oncol. 2001;19(5):1266-1274.

EuroQol-5D Utility Index (EQ-5D)

Kvam AKF, Peter M.; Wisloff, Finn. Responsiveness and minimal important score differences in quality-of-life questionnaires: a comparison of the EORTC QLQ-C30 cancer-specific questionnaire to the generic utility questionnaires EQ-5D and 15D in patients with multiple myeloma. Eur J Haematol. 2011;87(4):330-337.

EuroQol-5D Utility Index (EQ-5D) - UK weights

Simon ASN, M. P.; Cella, D. Estimation of minimally important differences in EQ-5D utility and VAS scores in cancer. Health and Quality of Life Outcomes. 2007;5(70).

EuroQol-5D Utility Index (EQ-5D) - US weights

Simon ASN, M. P.; Cella, D. Estimation of minimally important differences in EQ-5D utility and VAS scores in cancer. Health and Quality of Life Outcomes. 2007;5(70).

EuroQol-5D Utility Index-3 Level Version (EQ-5D-3L) - UK weights

Sagberg LMJ, A. S.; Solheim, O. Quality of life assessed with EQ-5D in patients undergoing glioma surgery: what is the responsiveness and minimal clinically important difference? Qual Life Res. 2014;23(5):1427-1434.

EuroQol-5D visual analogue scale (EQ-5D-VAS)

Simon ASN, M. P.; Cella, D. Estimation of minimally important differences in EQ-5D utility and VAS scores in cancer. Health and Quality of Life Outcomes. 2007;5(70).

Functional Assessment of Cancer Therapy (FACT) Advanced Prostate Symptom Index (FAPSI)

Cella DN, M. B.; Eton, D.; Nelson, J. B.; Mulani, P. Estimating clinically meaningful changes for the functional assessment of cancer therapy - Prostate: Results from a clinical trial of patients with metastatic hormone-refractory prostate cancer. Value Health. 2009;12(1):124-129.

Functional Assessment of Cancer Therapy (FACT) Head and Neck Symptom Index (FHNSI) (embedded)

Yount SL, Marcy; Du, Hongyan; Yost, Kathleen; Bode, Rita; Brockstein, Bruce; Argiris, Athanassios; Vokes, Everett; Cohen, Ezra E. W.; Campbell, Bruce; Valenzuela, Veronica; George, Jacquelyn; Egan, Robyn; Chen, Jessica; Meddis, David; Cella, David. A randomized validation study comparing embedded versus extracted FACT Head and Neck Symptom Index scores. Qual Life Res. 2007;16(10):1615-1626.

Functional Assessment of Cancer Therapy (FACT) Head and Neck Symptom Index (FHNSI) (stand-alone)

Yount SL, Marcy; Du, Hongyan; Yost, Kathleen; Bode, Rita; Brockstein, Bruce; Argiris, Athanassios; Vokes, Everett; Cohen, Ezra E. W.; Campbell, Bruce; Valenzuela, Veronica; George, Jacquelyn; Egan, Robyn; Chen, Jessica; Meddis, David; Cella, David. A randomized validation study comparing embedded versus extracted FACT Head and Neck Symptom Index scores. Qual Life Res. 2007;16(10):1615-1626.

Functional Assessment of Cancer Therapy (FACT) Trial Outcome Index-Anemia (TOI-An)

Cella DE, David T.; Lai, Jin-Shei; Peterman, Amy H.; Merkel, Douglas E. Combining anchor and distribution-based methods to derive minimal clinically important differences on the Functional Assessment of Cancer Therapy (FACT) anemia and fatigue scales. J Pain Symptom Manage. 2002;24(6):547-561.

Functional Assessment of Cancer Therapy (FACT) Trial Outcome Index-Fatigue (TOI-F)

Cella DE, David T.; Lai, Jin-Shei; Peterman, Amy H.; Merkel, Douglas E. Combining anchor and distribution-based methods to derive minimal clinically important differences on the Functional Assessment of Cancer Therapy (FACT) anemia and fatigue scales. J Pain Symptom Manage. 2002;24(6):547-561.

Functional Assessment of Cancer Therapy (FACT) Trial Outcome Index-Physical/Functional/Breast (TOI-PFB)

Eton DTC, David; Yost, Kathleen J.; Yount, Susan E.; Peterman, Amy H.; Neuberg, Donna S.; Sledge, George W.; Wood, William C. A combination of distribution- and anchor-based approaches determined minimally important differences (MIDs) for four endpoints in a breast cancer scale. J Clin Epidemiol. 2004;57(9):898-910.

Functional Assessment of Cancer Therapy (FACT) Trial Outcome Index–Colorectal (TOI-C)

Yost KJC, D.; Chawla, A.; Holmgren, E.; Eton, D. T.; Ayanian, J. Z.; West, D. W. Minimally important differences were estimated for the Functional Assessment of Cancer Therapy-Colorectal (FACT-C) instrument using a combination of distribution- and anchor-based approaches. J Clin Epidemiol. 2005;58(12):1241-1251.

Functional Assessment of Cancer Therapy (FACT)-Kidney Symptom Index (FKSI-10)

Cella DY, Susan; Du, Hongyan; Dhanda, Rahul; Gondek, Kathleen; Langefeld, Katie; George, Jacquelyn; Bro, William P.; Kelly, Celeste; Bukowski, Ronald. Development and validation of the Functional Assessment of Cancer Therapy-Kidney Symptom Index (FKSI). J Support Oncol. 2006;4(4):191-199.

Functional Assessment of Cancer Therapy (FACT)-Kidney Symptom Index (FKSI-15)

Cella DY, Susan; Du, Hongyan; Dhanda, Rahul; Gondek, Kathleen; Langefeld, Katie; George, Jacquelyn; Bro, William P.; Kelly, Celeste; Bukowski, Ronald. Development and validation of the Functional Assessment of Cancer Therapy-Kidney Symptom Index (FKSI). J Support Oncol. 2006;4(4):191-199.

Functional Assessment of Cancer Therapy (FACT)-Lung Symptom Index (FLSI-12)

Eton DTC, David; Yount, Susan E.; Davis, Kimberly M. Validation of the functional assessment of cancer therapy--lung symptom index-12 (FLSI-12). Lung Cancer. 2007;57(3):339-347.

Functional Assessment of Cancer Therapy-Anemia (FACT-An)

Cella DE, David T.; Lai, Jin-Shei; Peterman, Amy H.; Merkel, Douglas E. Combining anchor and distribution-based methods to derive minimal clinically important differences on the Functional Assessment of Cancer Therapy (FACT) anemia and fatigue scales. J Pain Symptom Manage. 2002;24(6):547-561.

Patrick DLG, D. D.; Zagari, M. J.; Mathijs, R.; Sweetenham, J.; Epoetin Alfa Study, Group. Assessing the clinical significance of health-related quality of life (HrQOL) improvements in anaemic cancer patients receiving epoetin alfa. Eur J Cancer. 2003;39(3):335-345.

Functional Assessment of Cancer Therapy-Breast Cancer (FACT-B)

Eton DTC, David; Yost, Kathleen J.; Yount, Susan E.; Peterman, Amy H.; Neuberg, Donna S.; Sledge, George W.; Wood, William C. A combination of distribution- and anchor-based approaches determined minimally important differences (MIDs) for four endpoints in a breast cancer scale. J Clin Epidemiol. 2004;57(9):898-910.

Functional Assessment of Cancer Therapy-Cognitive Function (FACT-Cog)

Cheung YTF, Y. L.; Shwe, M.; Tan, Y. P.; Fan, G.; Yong, W. S.; Madhukumar, P.; Ooi, W. S.; Chay, W. Y.; Dent, R. A.; Ang, S. F.; Lo, S. K.; Yap, Y. S.; Ng, R.; Chan, A. Minimal clinically important difference (MCID) for the functional assessment of cancer therapy: cognitive function (FACT-Cog) in breast cancer patients. J Clin Epidemiol. 2014;67(7):811-820.

Functional Assessment of Cancer Therapy-Colorectal cancer (FACT-C)

Yost KJC, D.; Chawla, A.; Holmgren, E.; Eton, D. T.; Ayanian, J. Z.; West, D. W. Minimally important differences were estimated for the Functional Assessment of Cancer Therapy-Colorectal (FACT-C) instrument using a combination of distribution- and anchor-based approaches. J Clin Epidemiol. 2005;58(12):1241-1251.

Functional Assessment of Cancer Therapy-Fatigue (FACT-F)

Cella DE, David T.; Lai, Jin-Shei; Peterman, Amy H.; Merkel, Douglas E. Combining anchor and distribution-based methods to derive minimal clinically important differences on the Functional Assessment of Cancer Therapy (FACT) anemia and fatigue scales. J Pain Symptom Manage. 2002;24(6):547-561.

Functional Assessment of Cancer Therapy-Gastric cancer (FACT-Ga)

Garland SNP, Guy; Lawe, Andrew; Biagioni, Bradly J.; Easaw, Jay; Eliasziw, Michael; Cella, David; Bathe, Oliver F. Prospective evaluation of the reliability, validity, and minimally important difference of the functional assessment of cancer therapy-gastric (FACT-Ga) quality-of-life instrument. Cancer. 2011;117(6):1302-1312.

Functional Assessment of Cancer Therapy-General (FACT-G)

Eton DTC, David; Yost, Kathleen J.; Yount, Susan E.; Peterman, Amy H.; Neuberg, Donna S.; Sledge, George W.; Wood, William C. A combination of distribution- and anchor-based approaches determined minimally important differences (MIDs) for four endpoints in a breast cancer scale. J Clin Epidemiol. 2004;57(9):898-910.

Cella DE, David T.; Lai, Jin-Shei; Peterman, Amy H.; Merkel, Douglas E. Combining anchor and distribution-based methods to derive minimal clinically important differences on the Functional Assessment of Cancer Therapy (FACT) anemia and fatigue scales. J Pain Symptom Manage. 2002;24(6):547-561.

Ringash JOS, Brian; Bezjak, Andrea; Redelmeier, Donald A. Interpreting clinically significant changes in patient-reported outcomes. Cancer. 2007;110(1):196-202.

Yount SL, Marcy; Du, Hongyan; Yost, Kathleen; Bode, Rita; Brockstein, Bruce; Argiris, Athanassios; Vokes, Everett; Cohen, Ezra E. W.; Campbell, Bruce; Valenzuela, Veronica; George, Jacquelyn; Egan, Robyn; Chen, Jessica; Meddis, David; Cella, David. A randomized validation study comparing embedded versus extracted FACT Head and Neck Symptom Index scores. Qual Life Res. 2007;16(10):1615-1626.

Patrick DLG, D. D.; Zagari, M. J.; Mathijs, R.; Sweetenham, J.; Epoetin Alfa Study, Group. Assessing the clinical significance of health-related quality of life (HrQOL) improvements in anaemic cancer patients receiving epoetin alfa. Eur J Cancer. 2003;39(3):335-345.

Cella DY, Susan; Du, Hongyan; Dhanda, Rahul; Gondek, Kathleen; Langefeld, Katie; George, Jacquelyn; Bro, William P.; Kelly, Celeste; Bukowski, Ronald. Development and validation of the Functional Assessment of Cancer Therapy-Kidney Symptom Index (FKSI). J Support Oncol. 2006;4(4):191-199.

Functional Assessment of Cancer Therapy-General (FACT-G) - version 3

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(3):207-221.

Functional Assessment of Cancer Therapy-General (FACT-H&N)

Yount SL, Marcy; Du, Hongyan; Yost, Kathleen; Bode, Rita; Brockstein, Bruce; Argiris, Athanassios; Vokes, Everett; Cohen, Ezra E. W.; Campbell, Bruce; Valenzuela, Veronica; George, Jacquelyn; Egan, Robyn; Chen, Jessica; Meddis, David; Cella, David. A randomized validation study comparing embedded versus extracted FACT Head and Neck Symptom Index scores. Qual Life Res. 2007;16(10):1615-1626.

Functional Assessment of Cancer Therapy-Head and Neck (FACT-H&N)

Ringash JOS, Brian; Bezjak, Andrea; Redelmeier, Donald A. Interpreting clinically significant changes in patient-reported outcomes. Cancer. 2007;110(1):196-202.

Ringash JB, Andrea; O'Sullivan, Brian; Redelmeier, Donald A. Interpreting differences in quality of life: the FACT-H&N in laryngeal cancer patients. Qual Life Res. 2004;13(4):725-733.

Functional Assessment of Cancer Therapy-Melanoma (FACT-M)

Askew RLX, Yan; Palmer, J. Lynn; Cella, David; Moye, Lemuel A.; Cormier, Janice N. Evaluating minimal important differences for the FACT-Melanoma quality of life questionnaire. Value Health. 2009;12(8):1144-1150.

Functional Assessment of Cancer Therapy-Prostate cancer (FACT-P)

Cella DN, M. B.; Eton, D.; Nelson, J. B.; Mulani, P. Estimating clinically meaningful changes for the functional assessment of cancer therapy - Prostate: Results from a clinical trial of patients with metastatic hormone-refractory prostate cancer. Value Health. 2009;12(1):124-129.

Functional Assessment of Cancer Therapy-Prostate cancer Trial Outcome Index (FACT-P TOI)

Cella DN, M. B.; Eton, D.; Nelson, J. B.; Mulani, P. Estimating clinically meaningful changes for the functional assessment of cancer therapy - Prostate: Results from a clinical trial of patients with metastatic hormone-refractory prostate cancer. Value Health. 2009;12(1):124-129.

Functional Assessment of Cancer Therapy/National Comprehensive Cancer Network (NCCN-FACT) Colorectal Cancer Symptom Index (FCSI-9)

Colwell HHM, Susan D.; Turner, Michelle P.; Lu, John; Wright, Nicola; Peeters, Marc; Cella, David; Devercelli, Giovanna. Psychometric evaluation of the FACT Colorectal Cancer Symptom Index (FCSI-9): reliability, validity, responsiveness, and clinical meaningfulness. Oncologist. 2010;15(3):308-316.

Functional Assessment of Cancer Treatment-General (FACT-G)

Garland SNP, Guy; Lawe, Andrew; Biagioni, Bradly J.; Easaw, Jay; Eliasziw, Michael; Cella, David; Bathe, Oliver F. Prospective evaluation of the reliability, validity, and minimally important difference of the functional assessment of cancer therapy-gastric (FACT-Ga) quality-of-life instrument. Cancer. 2011;117(6):1302-1312.

Karnofsky performance status (KPS)

Ringash JOS, Brian; Bezjak, Andrea; Redelmeier, Donald A. Interpreting clinically significant changes in patient-reported outcomes. Cancer. 2007;110(1):196-202.

Modified Health Assessment Questionnaire (MHAQ) Disability Index

Purcell AF, Jennifer; Bennett, Sally; Burmeister, Bryan; Haines, Terry. Determining the minimal clinically important difference criteria for the Multidimensional Fatigue Inventory in a radiotherapy population. Support Care Cancer. 2010;18(3):307-315.

Multidimensional Fatigue Inventory (MFI)

Purcell AF, Jennifer; Bennett, Sally; Burmeister, Bryan; Haines, Terry. Determining the minimal clinically important difference criteria for the Multidimensional Fatigue Inventory in a radiotherapy population. Support Care Cancer. 2010;18(3):307-315.

Perform Questionnaire (PQ)

Baro EC, Joan; Cassinello, Javier; Colomer, Ramon; Mata, Jesus Garcia; Gascon, Pere; Gasquet, Jose Antonio; Rodriguez, Cesar A.; Valentin, Vicente. Psychometric properties of the Perform Questionnaire: a brief scale for assessing patient perceptions of fatigue in cancer. Support Care Cancer. 2011;19(5):657-666.

PROMIS-Cancer Anxiety (Anxiety-9)

Yost KJE, David T.; Garcia, Sofia F.; Cella, David. Minimally important differences were estimated for six Patient-Reported Outcomes Measurement Information System-Cancer scales in advanced-stage cancer patients. J Clin Epidemiol. 2011;64(5):507-516.

PROMIS-Cancer Depression (Depression-10)

Yost KJE, David T.; Garcia, Sofia F.; Cella, David. Minimally important differences were estimated for six Patient-Reported Outcomes Measurement Information System-Cancer scales in advanced-stage cancer patients. J Clin Epidemiol. 2011;64(5):507-516.

PROMIS-Cancer Fatigue (Fatigue-17)

Yost KJE, David T.; Garcia, Sofia F.; Cella, David. Minimally important differences were estimated for six Patient-Reported Outcomes Measurement Information System-Cancer scales in advanced-stage cancer patients. J Clin Epidemiol. 2011;64(5):507-516.

PROMIS-Cancer Fatigue (Fatigue-7)

Yost KJE, David T.; Garcia, Sofia F.; Cella, David. Minimally important differences were estimated for six Patient-Reported Outcomes Measurement Information System-Cancer scales in advanced-stage cancer patients. J Clin Epidemiol. 2011;64(5):507-516.

PROMIS-Cancer Pain Interference (PainInt-10)

Yost KJE, David T.; Garcia, Sofia F.; Cella, David. Minimally important differences were estimated for six Patient-Reported Outcomes Measurement Information System-Cancer scales in advanced-stage cancer patients. J Clin Epidemiol. 2011;64(5):507-516.

PROMIS-Cancer Physical Function (PhysFunc-10)

Yost KJE, David T.; Garcia, Sofia F.; Cella, David. Minimally important differences were estimated for six Patient-Reported Outcomes Measurement Information System-Cancer scales in advanced-stage cancer patients. J Clin Epidemiol. 2011;64(5):507-516.

Short Form Health Survey 36-Item (SF-36)

Patrick DLG, D. D.; Zagari, M. J.; Mathijs, R.; Sweetenham, J.; Epoetin Alfa Study, Group. Assessing the clinical significance of health-related quality of life (HrQOL) improvements in anaemic cancer patients receiving epoetin alfa. Eur J Cancer. 2003;39(3):335-345.

Jayadevappa RM, S. B.; Wittink, M.; Wein, A. J.; Chhatre, S. Comparison of distribution- and anchor-based approaches to infer changes in health-related quality of life of prostate cancer survivors. Health Serv Res. 2012;47(5):1902-1925.

Social Difficulties Inventory (SDI)

Wright PM, Laura; Smith, Adam B.; Velikova, Galina; Selby, Peter. Measurement and interpretation of social distress using the social difficulties inventory (SDI). Eur J Cancer. 2008;44(11):1529-1535.

The Expanded Prostate Cancer Index Composite-Short Form 26 (EPIC-26)

Skolarus TAD, R. L.; Sanda, M. G.; Chang, P.; Greenfield, T. K.; Litwin, M. S.; Wei, J. T. Minimally important difference for the expanded prostate cancer index composite short form. Urology. 2015;85(1):101-105.

Time Trade Off (TTO)

Ringash JOS, Brian; Bezjak, Andrea; Redelmeier, Donald A. Interpreting clinically significant changes in patient-reported outcomes. Cancer. 2007;110(1):196-202.

UCLA prostate cancer index (UCLA-PCI)

Jayadevappa RM, S. B.; Wittink, M.; Wein, A. J.; Chhatre, S. Comparison of distribution- and anchor-based approaches to infer changes in health-related quality of life of prostate cancer survivors. Health Serv Res. 2012;47(5):1902-1925.

Work Limitations Questionnaire (WLQ)

Tamminga SJV, J. H.; Frings-Dresen, M. H.; De Boer, A. G. Measurement properties of the Work Limitations Questionnaire were sufficient among cancer survivors. Qual Life Res. 2014;23(2):515-525.

World Health Organization Quality Of Life Assessment Instrument (WHOQOL-100)

Den Oudsten BLZ, W. P.; De Vries, J. The minimal clinical important difference in the World Health Organization Quality of Life instrument - 100. Support Care Cancer. 2013;21(5):1295-1301.

Ophthalmology

Graves' Ophthalmopathy Quality of Life Questionnaire (GO-QOL)

Terwee CBD, F. W.; Mourits, M. P.; Gerding, M. N.; Baldeschi, L.; Kalmann, R.; Prummel, M. F.; Wiersinga, W. M. Interpretation and validity of changes in scores on the Graves' ophthalmopathy quality of life questionnaire (GO-QOL) after different treatments. Clin Endocrinol (Oxf). 2001;54(3):391-398.

Impact Of Dry Eye On Everyday Life (IDEEL)

Fairchild CJC, Robin L.; Begley, Carolyn G. Clinically important difference in dry eye: change in IDEEL-symptom bother. Optom Vis Sci. 2008;85(8):699-707.

Modified Low Vision Quality of Life questionnaire (LVQOL)

de Boer MRdV, Henrica C. W.; Terwee, Caroline B.; Moll, Annette C.; Volker-Dieben, Hennie J. M.; van Rens, Ger H. M. B. Changes to the subscales of two vision-related quality of life questionnaires are proposed. J Clin Epidemiol. 2005;58(12):1260-1268.

Modified Vision-Related Quality of Life Core Measure (VCM1)

de Boer MRdV, Henrica C. W.; Terwee, Caroline B.; Moll, Annette C.; Volker-Dieben, Hennie J. M.; van Rens, Ger H. M. B. Changes to the subscales of two vision-related quality of life questionnaires are proposed. J Clin Epidemiol. 2005;58(12):1260-1268.

National Eye Institute-Visual Function Questionnaire-25 (NEI-VFQ-25)

Gillespie BWM, D. C.; Niziol, L. M.; Janz, N. K. Estimating minimally important differences for two vision-specific quality of life measures. Invest Ophthalmol Vis Sci. 2014;55(7):4206-4212.

Naik RKG, K. S.; Rentz, A. M.; Kowalski, J. W.; Revicki, D. A. Psychometric evaluation of the National Eye Institute Visual Function Questionnaire and Visual Function Questionnaire Utility Index in patients with non-infectious intermediate and posterior uveitis. Qual Life Res. 2013;22(10):2801-2808.

Ocular Surface Disease Index (OSDI)

Miller KLW, John G.; Mink, David R.; Satram-Hoang, Sacha; Wilson, Steven E.; Perry, Henry D.; Asbell, Penny A.; Pflugfelder, Stephen C. Minimal clinically important difference for the ocular surface disease index. Arch Ophthalmol. 2010;128(1):94-101.

Short Form Health Survey 36-Item (SF-36)

Bilbao AQ, Jose M.; Escobar, Antonio; Garcia, Susana; Andradas, Elena; Bare, Marisa; Elizalde, Belen; Group, I. RYSS-Cataract. Responsiveness and clinically important differences for the VF-14 index, SF-36, and visual acuity in patients undergoing cataract surgery. Ophthalmology. 2009;116(3):418-424.e411.

Visual Activities Questionnaire (VAQ)

Gillespie BWM, D. C.; Niziol, L. M.; Janz, N. K. Estimating minimally important differences for two vision-specific quality of life measures. Invest Ophthalmol Vis Sci. 2014;55(7):4206-4212.

Visual Function Index (VF-14)

Browne JPvdM, Jan H.; Lewsey, James D.; Lamping, Donna L.; Black, Nick. Mathematical coupling may account for the association between baseline severity and minimally important difference values. J Clin Epidemiol. 2010;63(8):865-874.

Bilbao AQ, Jose M.; Escobar, Antonio; Garcia, Susana; Andradas, Elena; Bare, Marisa; Elizalde, Belen; Group, I. RYSS-Cataract. Responsiveness and clinically important differences for the VF-14 index, SF-36, and visual acuity in patients undergoing cataract surgery. Ophthalmology. 2009;116(3):418-424.e411.

Quintana JMA, Urko; Las-Hayas, Carlota; Gonzalez, Nerea; Garcia, Susana; Escobar, Antonio; Investigacion En Resultados, Y. Servicios De Salud Cataract Group. Use of the patient acceptable symptom state and the minimal clinically important difference to evaluate the outcomes of cataract extraction. Am J Ophthalmol. 2011;152(2):234-243.e233.

Quintana JME, Antonio; Bilbao, Amaia; Blasco, Juan A.; Lacalle, Juan R.; Bare, Marisa; Begiristain, Jose M.; Group, I. RYSS-Cataract. Validity of newly developed appropriateness criteria for cataract surgery. Ophthalmology. 2009;116(3):409-417.e403.

Visual Function Questionnaire Utility Index (VFQ-UI)

Naik RKG, K. S.; Rentz, A. M.; Kowalski, J. W.; Revicki, D. A. Psychometric evaluation of the National Eye Institute Visual Function Questionnaire and Visual Function Questionnaire Utility Index in patients with non-infectious intermediate and posterior uveitis. Qual Life Res. 2013;22(10):2801-2808.

Orthopedic Surgery

American Shoulder and Elbow Surgeons Standardized Shoulder Assessment Form (ASES)

Michener LAM, Philip W.; Sennett, Brian J. American Shoulder and Elbow Surgeons Standardized Shoulder Assessment Form, patient self-report section: reliability, validity, and responsiveness. J Shoulder Elbow Surg. 2002;11(6):587-594.

Tashjian RZD, Julia; Green, Andrew; Porucznik, Christina A.; Powell, Amy P. Minimal clinically important differences in ASES and simple shoulder test scores after nonoperative treatment of rotator cuff disease. J Bone Joint Surg Am. 2010;92(2):296-303.

Copenhagen Hip and Groin Outcome Score (HAGOS)

Kemp JLC, N. J.; Roos, E. M.; Crossley, K. M. Psychometric properties of patient-reported outcome measures for hip arthroscopic surgery. Am J Sports Med. 2013;41(9):2065-2073.

Disabilities of the Arm, Shoulder and Hand (DASH) - Danish version

Lundquist CBD, K.; Christiansen, D. H. Responsiveness of a Danish version of the Disabilities of the Arm, Shoulder and Hand (DASH) questionnaire. Dan Med J. 2014;61(4):A4813.

Disabilities of the Arm, Shoulder, and Hand (DASH)

Sorensen AAH, D.; Tan, W. H.; Ketchersid, J.; Calfee, R. P. Minimal clinically important differences of 3 patient-rated outcomes instruments. J Hand Surg [Am]. 2013;38(4):641-649.

Dawson JD, Helen; Boller, Irene; Fitzpatrick, Ray; Little, Christopher; Rees, Jonathan; Carr, Andrew. Comparative responsiveness and minimal change for the Oxford Elbow Score following surgery. Qual Life Res. 2008;17(10):1257-1267.

Beaton DEvE, Dwayne; Smith, Peter; van der Velde, Gabrielle; Cullen, Kimberley; Kennedy, Carol A.; Hogg-Johnson, Sheilah. Minimal change is sensitive, less specific to recovery: a diagnostic testing approach to interpretability. J Clin Epidemiol. 2011;64(5):487-496.

Schmitt JSDF, Richard P. Reliable change and minimum important difference (MID) proportions facilitated group responsiveness comparisons using individual threshold criteria. J Clin Epidemiol. 2004;57(10):1008-1018.

van Kampen DAW, W. J.; van Beers, L. W.; Castelein, R. M.; Scholtes, V. A.; Terwee, C. B. Determination and comparison of the smallest detectable change (SDC) and the minimal important change (MIC) of four-shoulder patient-reported outcome measures (PROMs). Journal of Orthopaedic Surgery. 2013;8:40.

van de Water ATMS, Nora; Davidson, Megan; Evans, Matthew; Taylor, Nicholas F. Reliability and validity of shoulder function outcome measures in people with a proximal humeral fracture. Disabil Rehabil. 2014;36(13):1072-1079.

Disabilities of the Arm, Shoulder, and Hand (QuickDASH)

Sorensen AAH, D.; Tan, W. H.; Ketchersid, J.; Calfee, R. P. Minimal clinically important differences of 3 patient-rated outcomes instruments. J Hand Surg [Am]. 2013;38(4):641-649.

van Kampen DAW, W. J.; van Beers, L. W.; Castelein, R. M.; Scholtes, V. A.; Terwee, C. B. Determination and comparison of the smallest detectable change (SDC) and the minimal important change (MIC) of four-shoulder patient-reported outcome measures (PROMs). Journal of Orthopaedic Surgery. 2013;8:40.

EuroQol-5D Utility Index (EQ-5D)

Walters SJB, John E. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. Qual Life Res. 2005;14(6):1523-1532.

Browne JPvdM, Jan H.; Lewsey, James D.; Lamping, Donna L.; Black, Nick. Mathematical coupling may account for the association between baseline severity and minimally important difference values. J Clin Epidemiol. 2010;63(8):865-874.

Impellizzeri FMM, A. F.; Naal, F. D.; Hersche, O.; Leunig, M. The early outcome of surgical treatment for femoroacetabular impingement: success depends on how you measure it. Osteoarthritis Cartilage. 2012;20(7):638-645.

Paulsen AR, E. M.; Pedersen, A. B.; Overgaard, S. Minimal clinically important improvement (MCII) and patient-acceptable symptom state (PASS) in total hip arthroplasty (THA) patients 1 year postoperatively. Acta Orthopaedica. 2014;85(1):39-48.

EuroQol-5D visual analogue scale (EQ-5D-VAS)

Impellizzeri FMM, A. F.; Naal, F. D.; Hersche, O.; Leunig, M. The early outcome of surgical treatment for femoroacetabular impingement: success depends on how you measure it. Osteoarthritis Cartilage. 2012;20(7):638-645.

Paulsen AR, E. M.; Pedersen, A. B.; Overgaard, S. Minimal clinically important improvement (MCII) and patient-acceptable symptom state (PASS) in total hip arthroplasty (THA) patients 1 year postoperatively. Acta Orthopaedica. 2014;85(1):39-48.

Hip Disability and Osteoarthritis Outcome Score-Physical Function Shortform (HOOS-PS)

Paulsen AR, E. M.; Pedersen, A. B.; Overgaard, S. Minimal clinically important improvement (MCII) and patient-acceptable symptom state (PASS) in total hip arthroplasty (THA) patients 1 year postoperatively. Acta Orthopaedica. 2014;85(1):39-48.

Hip Disability and Osteoarthritis Outcome Score (HOOS)

Paulsen AR, E. M.; Pedersen, A. B.; Overgaard, S. Minimal clinically important improvement (MCII) and patient-acceptable symptom state (PASS) in total hip arthroplasty (THA) patients 1 year postoperatively. Acta Orthopaedica. 2014;85(1):39-48.

Kemp JLC, N. J.; Roos, E. M.; Crossley, K. M. Psychometric properties of patient-reported outcome measures for hip arthroscopic surgery. Am J Sports Med. 2013;41(9):2065-2073.

Hip Outcome Score (HOS)

Kemp JLC, N. J.; Roos, E. M.; Crossley, K. M. Psychometric properties of patient-reported outcome measures for hip arthroscopic surgery. Am J Sports Med. 2013;41(9):2065-2073.

Intermittent and constant osteoarthritis pain (ICOAP)

Singh JAL, R.; Landon, G. C.; Suarez-Almazor, M. Reliability and clinically important improvement thresholds for osteoarthritis pain and function scales: a multicenter study. J Rheumatol. 2014;41(3):509-515.

International Hip Outcome Tool (iHOT-33)

Kemp JLC, N. J.; Roos, E. M.; Crossley, K. M. Psychometric properties of patient-reported outcome measures for hip arthroscopic surgery. Am J Sports Med. 2013;41(9):2065-2073.

International Knee Documentation Committee (IKDC) Subjective Knee Form

Greco NJA, Allen F.; Mann, Barton J.; Cole, Brian J.; Farr, Jack; Nissen, Carl W.; Irrgang, James J. Responsiveness of the International Knee Documentation Committee Subjective Knee Form in comparison to the Western Ontario and McMaster Universities Osteoarthritis Index, modified Cincinnati Knee Rating System, and Short Form 36 in patients with focal articular cartilage defects. Am J Sports Med. 2010;38(5):891-902.

Knee injury and Osteoarthritis Outcome Score (KOOS)

Singh JAL, R.; Landon, G. C.; Suarez-Almazor, M. Reliability and clinically important improvement thresholds for osteoarthritis pain and function scales: a multicenter study. J Rheumatol. 2014;41(3):509-515.

Knee Injury and Osteoarthritis Outcome Score (KOOS) - Italian version

Monticone MF, S.; Salvaderi, S.; Motta, L.; Cerri, C. Responsiveness and minimal important changes for the Knee Injury and Osteoarthritis Outcome Score in subjects undergoing rehabilitation after total knee arthroplasty. Am J Phys Med Rehabil. 2013;92(10):864-870.

Knee injury and Osteoarthritis Outcome Score-Physical Function Shortform (KOOS-PS)

Singh JAL, R.; Landon, G. C.; Suarez-Almazor, M. Reliability and clinically important improvement thresholds for osteoarthritis pain and function scales: a multicenter study. J Rheumatol. 2014;41(3):509-515.

Knee Quality of Life (KQoL-26)

Chuang LHG, A.; Brealey, S. Comparative responsiveness and minimal change of the Knee Quality of Life 26-item (KQoL-26) questionnaire. Qual Life Res. 2013;22(9):2461-2475.

Lower Extremity Functional Scale (LEFS) using computerized adaptive test (CAT)

Hart DL, Wang YC, Stratford PW, Mioduski JE. A computerized adaptive test for patients with hip impairments produced valid and responsive measures of function. Archives of physical medicine and rehabilitation. 2008;89(11):2129-2139.

Lysholm Knee Score

Chuang LHG, A.; Brealey, S. Comparative responsiveness and minimal change of the Knee Quality of Life 26-item (KQoL-26) questionnaire. Qual Life Res. 2013;22(9):2461-2475.

Manchester-Oxford foot questionnaire (MOxFQ)

Dawson JD, H.; Coffey, J.; Jenkinson, C.; Oxford; Birmingham, Foot; Ankle Clinical Research, Group. Responsiveness and minimally important change for the Manchester-Oxford foot questionnaire (MOXFQ) compared with AOFAS and SF-36 assessments following surgery for hallux valgus.[Erratum appears in Osteoarthritis Cartilage. 2011 Jul 19(7):920]. Osteoarthritis Cartilage. 2007;15(8):918-931.

Dawson JB, I.; Doll, H.; Lavis, G.; Sharp, R.; Cooke, P.; Jenkinson, C. Minimally important change was estimated for the Manchester-Oxford Foot Questionnaire after foot/ankle surgery. J Clin Epidemiol. 2014;67(6):697-705.

Michigan Hand Outcomes Questionnaire (MHQ)

London DAS, J. G.; Calfee, R. P. Determining the Michigan Hand Outcomes Questionnaire minimal clinically important difference by means of three methods. Plast Reconstr Surg. 2014;133(3):616-625.

Modified Cincinnati Knee Rating System (CKRS)

Greco NJA, Allen F.; Mann, Barton J.; Cole, Brian J.; Farr, Jack; Nissen, Carl W.; Irrgang, James J. Responsiveness of the International Knee Documentation Committee Subjective Knee Form in comparison to the Western Ontario and McMaster Universities Osteoarthritis Index, modified Cincinnati Knee Rating System, and Short Form 36 in patients with focal articular cartilage defects. Am J Sports Med. 2010;38(5):891-902.

Modified Harris Hip Score (MHHS)

Kemp JLC, N. J.; Roos, E. M.; Crossley, K. M. Psychometric properties of patient-reported outcome measures for hip arthroscopic surgery. Am J Sports Med. 2013;41(9):2065-2073.

Oxford Elbow Score (OES)

Dawson JD, Helen; Boller, Irene; Fitzpatrick, Ray; Little, Christopher; Rees, Jonathan; Carr, Andrew. Comparative responsiveness and minimal change for the Oxford Elbow Score following surgery. Qual Life Res. 2008;17(10):1257-1267.

Oxford Hip Score (OHS)

Browne JPvdM, Jan H.; Lewsey, James D.; Lamping, Donna L.; Black, Nick. Mathematical coupling may account for the association between baseline severity and minimally important difference values. J Clin Epidemiol. 2010;63(8):865-874.

Impellizzeri FMM, A. F.; Naal, F. D.; Hersche, O.; Leunig, M. The early outcome of surgical treatment for femoroacetabular impingement: success depends on how you measure it. Osteoarthritis Cartilage. 2012;20(7):638-645.

Beard DJH, K.; Dawson, J.; Doll, H.; Murray, D. W.; Carr, A. J.; Price, A. J. Meaningful changes for the Oxford hip and knee scores after joint replacement surgery. J Clin Epidemiol. 2015;68(1):73-79.

Oxford Knee Score (OKS)

Browne JPvdM, Jan H.; Lewsey, James D.; Lamping, Donna L.; Black, Nick. Mathematical coupling may account for the association between baseline severity and minimally important difference values. J Clin Epidemiol. 2010;63(8):865-874.

Clement NDM, D.; Simpson, A. H. The minimal clinically important difference in the Oxford knee score and Short Form 12 score after total knee arthroplasty. Knee Surgery, Sports Traumatology, Arthroscopy. 2014;22(8):1933-1939.

Beard DJH, K.; Dawson, J.; Doll, H.; Murray, D. W.; Carr, A. J.; Price, A. J. Meaningful changes for the Oxford hip and knee scores after joint replacement surgery. J Clin Epidemiol. 2015;68(1):73-79.

Oxford Shoulder Score (OSS)

Ekeberg OMB-H, Erik; Keller, Anne; Tveita, Einar K.; Juel, Niels G.; Brox, Jens I. A questionnaire found disease-specific WORC index is not more responsive than SPADI and OSS in rotator cuff disease. J Clin Epidemiol. 2010;63(5):575-584.

van Kampen DAW, W. J.; van Beers, L. W.; Castelein, R. M.; Scholtes, V. A.; Terwee, C. B. Determination and comparison of the smallest detectable change (SDC) and the minimal important change (MIC) of four-shoulder patient-reported outcome measures (PROMs). Journal of Orthopaedic Surgery. 2013;8:40.

van de Water ATMS, Nora; Davidson, Megan; Evans, Matthew; Taylor, Nicholas F. Reliability and validity of shoulder function outcome measures in people with a proximal humeral fracture. Disabil Rehabil. 2014;36(13):1072-1079.

Patient-Rated Wrist Evaluation (PRWE)

Sorensen AAH, D.; Tan, W. H.; Ketchersid, J.; Calfee, R. P. Minimal clinically important differences of 3 patient-rated outcomes instruments. J Hand Surg [Am]. 2013;38(4):641-649.

Schmitt JSDF, Richard P. Reliable change and minimum important difference (MID) proportions facilitated group responsiveness comparisons using individual threshold criteria. J Clin Epidemiol. 2004;57(10):1008-1018.

Short Form Health Survey 12-Item (SF-12)

Schmitt JSDF, Richard P. Reliable change and minimum important difference (MID) proportions facilitated group responsiveness comparisons using individual threshold criteria. J Clin Epidemiol. 2004;57(10):1008-1018.

Clement NDM, D.; Simpson, A. H. The minimal clinically important difference in the Oxford knee score and Short Form 12 score after total knee arthroplasty. Knee Surgery, Sports Traumatology, Arthroscopy. 2014;22(8):1933-1939.

Short Form Health Survey 36-Item (SF-36)

Quintana JME, A.; Bilbao, A.; Arostegui, I.; Lafuente, I.; Vidaurreta, I. Responsiveness and clinically important differences for the WOMAC and SF-36 after hip joint replacement. Osteoarthritis Cartilage. 2005;13(12):1076-1083.

Dawson JD, H.; Coffey, J.; Jenkinson, C.; Oxford; Birmingham, Foot; Ankle Clinical Research, Group. Responsiveness and minimally important change for the Manchester-Oxford foot questionnaire (MOXFQ) compared with AOFAS and SF-36 assessments following surgery for hallux valgus.[Erratum appears in Osteoarthritis Cartilage. 2011 Jul 19(7):920]. Osteoarthritis Cartilage. 2007;15(8):918-931.

Greco NJA, Allen F.; Mann, Barton J.; Cole, Brian J.; Farr, Jack; Nissen, Carl W.; Irrgang, James J. Responsiveness of the International Knee Documentation Committee Subjective Knee Form in comparison to the Western Ontario and McMaster Universities Osteoarthritis Index, modified Cincinnati Knee Rating System, and Short Form 36 in patients with focal articular cartilage defects. Am J Sports Med. 2010;38(5):891-902.

Chuang LHG, A.; Brealey, S. Comparative responsiveness and minimal change of the Knee Quality of Life 26-item (KQoL-26) questionnaire. Qual Life Res. 2013;22(9):2461-2475.

Short Form Health Survey 36-Item (SF-36) - version 2

Dawson JB, I.; Doll, H.; Lavis, G.; Sharp, R.; Cooke, P.; Jenkinson, C. Minimally important change was estimated for the Manchester-Oxford Foot Questionnaire after foot/ankle surgery. J Clin Epidemiol. 2014;67(6):697-705.

Short-Form Six-Dimension (SF-6D)

Walters SJB, John E. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. Qual Life Res. 2005;14(6):1523-1532.

Quintana JME, A.; Bilbao, A.; Arostegui, I.; Lafuente, I.; Vidaurreta, I. Responsiveness and clinically important differences for the WOMAC and SF-36 after hip joint replacement. Osteoarthritis Cartilage. 2005;13(12):1076-1083.

Shoulder Pain and Disability Index (SPADI)

Ekeberg OMB-H, Erik; Keller, Anne; Tveita, Einar K.; Juel, Niels G.; Brox, Jens I. A questionnaire found disease-specific WORC index is not more responsive than SPADI and OSS in rotator cuff disease. J Clin Epidemiol. 2010;63(5):575-584.

Schmitt JSDF, Richard P. Reliable change and minimum important difference (MID) proportions facilitated group responsiveness comparisons using individual threshold criteria. J Clin Epidemiol. 2004;57(10):1008-1018.

Shoulder pain visual analogue scale (VAS)

Tashjian RZD, Julia; Porucznik, Christina A.; Powell, Amy P. Minimal clinically important differences (MCID) and patient acceptable symptomatic state (PASS) for visual analog scales (VAS) measuring pain in patients treated for rotator cuff disease. J Shoulder Elbow Surg. 2009;18(6):927-932.

Simple Shoulder Test (SST)

Tashjian RZD, Julia; Green, Andrew; Porucznik, Christina A.; Powell, Amy P. Minimal clinically important differences in ASES and simple shoulder test scores after nonoperative treatment of rotator cuff disease. J Bone Joint Surg Am. 2010;92(2):296-303.

van Kampen DAW, W. J.; van Beers, L. W.; Castelein, R. M.; Scholtes, V. A.; Terwee, C. B. Determination and comparison of the smallest detectable change (SDC) and the minimal important change (MIC) of four-shoulder patient-reported outcome measures (PROMs). Journal of Orthopaedic Surgery. 2013;8:40.

Subjective Shoulder Value (SSV)

van de Water ATMS, Nora; Davidson, Megan; Evans, Matthew; Taylor, Nicholas F. Reliability and validity of shoulder function outcome measures in people with a proximal humeral fracture. Disabil Rehabil. 2014;36(13):1072-1079.

Victorian Institute of Sport Assessment-Patellar Tendon (VISA-P) Questionnaire

Hernandez-Sanchez SH, M. D.; Gomez, A. Responsiveness of the VISA-P scale for patellar tendinopathy in athletes. BJSM online. 2014;48(6):453-457.

Victorian Institute of Sport Assessment-Proximal Hamstring Tendons (VISA-H) questionnaire

Cacchio ADP, F.; Maffulli, N. Development and validation of a new visa questionnaire (VISA-H) for patients with proximal hamstring tendinopathy. BJSM online. 2014;48(6):448-452.

Western Ontario and McMaster Universities Arthritis Index (WOMAC)

Impellizzeri FMM, A. F.; Naal, F. D.; Hersche, O.; Leunig, M. The early outcome of surgical treatment for femoroacetabular impingement: success depends on how you measure it. Osteoarthritis Cartilage. 2012;20(7):638-645.

Tubach FD, Maxime; Falissard, Bruno; Baron, Gabriel; Logeart, Isabelle; Ravaud, Philippe. Feeling good rather than feeling better matters more to patients. Arthritis Rheum. 2006;55(4):526-530.

Terwee CBR, Leo D.; Dekker, Joost; Bierma-Zeinstra, Sita M.; Peat, George; Jordan, Kelvin P.; Croft, Peter; de Vet, Henrica C. W. Mind the MIC: large variation among populations and methods. J Clin Epidemiol. 2010;63(5):524-534.

Quintana JMA, Urko; Barrio, Irantzu; Orive, Miren; Garcia, Susana; Escobar, Antonio. Outcomes after total hip replacement based on patients' baseline status: what results can be expected? Arthritis Care Res (Hoboken). 2012;64(4):563-572.

Quintana JME, A.; Bilbao, A.; Arostegui, I.; Lafuente, I.; Vidaurreta, I. Responsiveness and clinically important differences for the WOMAC and SF-36 after hip joint replacement. Osteoarthritis Cartilage. 2005;13(12):1076-1083.

Greco NJA, Allen F.; Mann, Barton J.; Cole, Brian J.; Farr, Jack; Nissen, Carl W.; Irrgang, James J. Responsiveness of the International Knee Documentation Committee Subjective Knee Form in comparison to the Western Ontario and McMaster Universities Osteoarthritis Index, modified Cincinnati Knee Rating System, and Short Form 36 in patients with focal articular cartilage defects. Am J Sports Med. 2010;38(5):891-902.

Escobar AGP, L.; Herrera-Espineira, C.; Aizpuru, F.; Sarasqueta, C.; Gonzalez Saenz de Tejada, M.; Quintana, J. M.; Bilbao, A. Total knee replacement; minimal clinically important differences and responders. Osteoarthritis Cartilage. 2013;21(12):2006-2012.

Western Ontario and McMaster University Arthritis Index (WOMAC) - Dutch version

Terwee CBR, Leo D.; Knol, Dirk L.; De Boer, Michiel R.; De Vet, Henrica C. W. Linking measurement error to minimal important change of patient-reported outcomes. J Clin Epidemiol. 2009;62(10):1062-1067.

Western Ontario Rotator Cuff Index (WORC)

Ekeberg OMB-H, Erik; Keller, Anne; Tveita, Einar K.; Juel, Niels G.; Brox, Jens I. A questionnaire found disease-specific WORC index is not more responsive than SPADI and OSS in rotator cuff disease. J Clin Epidemiol. 2010;63(5):575-584.

Orthopedic Surgery, Plastic Surgery

6-item Carpal Tunnel Symptoms Scale (CTS-6)

Atroshi IL, Per-Erik; Ornstein, Ewald; Gummesson, Christina. The six-item CTS symptoms scale and palmar pain scale in carpal tunnel syndrome. J Hand Surg [Am]. 2011;36(5):788-794.

Brigham and Women's Hospital Carpal Tunnel Syndrome Questionnaire or Boston Carpal Tunnel Syndrome Questionnaire (BCTSQ)

Ozyurekoglu TM, Steven J.; Goldsmith, L. Jane; LaJoie, A. Scott. The minimal clinically important difference of the Carpal Tunnel Syndrome Symptom Severity Scale. J Hand Surg [Am]. 2006;31(5):733-738; discussion 739-740.

Malay SSUNSGC, K. C. The minimal clinically important difference after simple decompression for ulnar neuropathy at the elbow. J Hand Surg [Am]. 2013;38(4):652-659.

Ozer KM, S.; Toker, S.; Chung, K. C. Minimal clinically important difference of carpal tunnel release in diabetic and nondiabetic patients. Plast Reconstr Surg. 2013;131(6):1279-1285.

Kim JKJ, S. H. Minimal clinically important differences in the Carpal Tunnel Questionnaire after carpal tunnel release. J. 2013;38(1):75-79.

Disabilities of the Arm, Shoulder, and Hand (DASH)

Kim JKP, E. S. Comparative responsiveness and minimal clinically important differences for idiopathic ulnar impaction syndrome. Clin Orthop. 2013;471(5):1406-1411.

Marks MA, L.; Herren, D. B.; Schindele, S.; Nelissen, R. G.; Vliet Vlieland, T. P. Measurement properties of the German Michigan Hand Outcomes Questionnaire in patients with trapeziometacarpal osteoarthritis. Arthritis Care Res (Hoboken). 2014;66(2):245-252.

Malay SSUNSGC, K. C. The minimal clinically important difference after simple decompression for ulnar neuropathy at the elbow. J Hand Surg [Am]. 2013;38(4):652-659.

Franchignoni FV, S.; Giordano, A.; Sartorio, F.; Bravini, E.; Ferriero, G. Minimal clinically important difference of the disabilities of the arm, shoulder and hand outcome measure (DASH) and its shortened version (QuickDASH). J Orthop Sports Phys Ther. 2014;44(1):30-39.

Disabilities of the Arm, Shoulder, and Hand (QuickDASH)

Franchignoni FV, S.; Giordano, A.; Sartorio, F.; Bravini, E.; Ferriero, G. Minimal clinically important difference of the disabilities of the arm, shoulder and hand outcome measure (DASH) and its shortened version (QuickDASH). J Orthop Sports Phys Ther. 2014;44(1):30-39.

Michigan Hand Outcomes Questionnaire (MHQ)

Shauver MJC, Kevin C. The minimal clinically important difference of the Michigan hand outcomes questionnaire. J Hand Surg [Am]. 2009;34(3):509-514.

Marks MA, L.; Herren, D. B.; Schindele, S.; Nelissen, R. G.; Vliet Vlieland, T. P. Measurement properties of the German Michigan Hand Outcomes Questionnaire in patients with trapeziometacarpal osteoarthritis. Arthritis Care Res (Hoboken). 2014;66(2):245-252.

Malay SSUNSGC, K. C. The minimal clinically important difference after simple decompression for ulnar neuropathy at the elbow. J Hand Surg [Am]. 2013;38(4):652-659.

Patient-Rated Wrist Evaluation (PRWE)

Kim JKP, E. S. Comparative responsiveness and minimal clinically important differences for idiopathic ulnar impaction syndrome. Clin Orthop. 2013;471(5):1406-1411.

Short Form Health Survey 12-Item (SF-12)

Marks MA, L.; Herren, D. B.; Schindele, S.; Nelissen, R. G.; Vliet Vlieland, T. P. Measurement properties of the German Michigan Hand Outcomes Questionnaire in patients with trapeziometacarpal osteoarthritis. Arthritis Care Res (Hoboken). 2014;66(2):245-252.

Orthopedic Surgery, Rheumatology

Disabilities of the Arm, Shoulder, and Hand (DASH)

Sorensen AAH, D.; Tan, W. H.; Ketchersid, J.; Calfee, R. P. Minimal clinically important differences of 3 patient-rated outcomes instruments. J Hand Surg [Am]. 2013;38(4):641-649.

Disabilities of the Arm, Shoulder, and Hand (QuickDASH)

Sorensen AAH, D.; Tan, W. H.; Ketchersid, J.; Calfee, R. P. Minimal clinically important differences of 3 patient-rated outcomes instruments. J Hand Surg [Am]. 2013;38(4):641-649.

Polson KR, Duncan; McNair, Peter J.; Larmer, Peter. Responsiveness, minimal importance difference and minimal detectable change scores of the shortened disability arm shoulder hand (QuickDASH) questionnaire. Manual Ther. 2010;15(4):404-407.

Disease activity numerical rating scale (NRS)

Tubach FR, P.; Martin-Mola, E.; Awada, H.; Bellamy, N.; Bombardier, C.; Felson, D. T.; Hajjaj-Hassouni, N.; Hochberg, M.; Logeart, I.; Matucci-Cerinic, M.; van de Laar, M.; van der Heijde, D.; Dougados, M. Minimum clinically important improvement and patient acceptable symptom state in pain and function in rheumatoid arthritis, ankylosing spondylitis, chronic back pain, hand osteoarthritis, and hip and knee osteoarthritis: Results from a prospective multinational study. Arthritis Care Res (Hoboken). 2012;64(11):1699-1707.

EuroQol-5D Utility Index (EQ-5D)

Walters SJB, John E. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. Qual Life Res. 2005;14(6):1523-1532.

Function numerical rating scale (NRS)

Ornetti PD, Maxime; Paternotte, Simon; Logeart, Isabelle; Gossec, Laure. Validation of a numerical rating scale to assess functional impairment in hip and knee osteoarthritis: comparison with the WOMAC function scale. Ann Rheum Dis. 2011;70(5):740-746.

Functional disability numerical rating scale (NRS)

Tubach FR, P.; Martin-Mola, E.; Awada, H.; Bellamy, N.; Bombardier, C.; Felson, D. T.; Hajjaj-Hassouni, N.; Hochberg, M.; Logeart, I.; Matucci-Cerinic, M.; van de Laar, M.; van der Heijde, D.; Dougados, M. Minimum clinically important improvement and patient acceptable symptom state in pain and function in rheumatoid arthritis, ankylosing spondylitis, chronic back pain, hand osteoarthritis, and hip and knee osteoarthritis: Results from a prospective multinational study. Arthritis Care Res (Hoboken). 2012;64(11):1699-1707.

Global assessment of disease activity visual analogue scale (VAS)

Tubach FR, P.; Baron, G.; Falissard, B.; Logeart, I.; Bellamy, N.; Bombardier, C.; Felson, D.; Hochberg, M.; van der Heijde, D.; Dougados, M. Evaluation of clinically relevant changes in patient reported outcomes in knee and hip osteoarthritis: the minimal clinically important improvement. Ann Rheum Dis. 2005;64(1):29-33.

Ibadan Knee/Hip Osteoarthritis Outcome Measure (IKHOAM)

Akinpelu AOO, A. C.; Raheem, S. Minimal clinically important difference of Ibadan knee hip osteoarthritis outcome measure (IKHOAM) a cross-sectional study. Journal of Physical Therapy. 2011;3(2):52-60.

Intermittent and Constant Osteoarthritis Pain (ICOAP)

Harris KKD, J.; Jones, L. D.; Beard, D. J.; Price, A. J. Extending the use of PROMs in the NHS-using the Oxford Knee Score in patients undergoing non-operative management for knee osteoarthritis: A validation study. BMJ Open. 2013;3(8).

Knee injury and Osteoarthritis Outcome Score-Physical Function Shortform (KOOS-PS)

Harris KKD, J.; Jones, L. D.; Beard, D. J.; Price, A. J. Extending the use of PROMs in the NHS-using the Oxford Knee Score in patients undergoing non-operative management for knee osteoarthritis: A validation study. BMJ Open. 2013;3(8).

Lower-Limb Tasks Questionnaire (LLTQ)

McNair PJP, Harry; Collier, Jill; Bassett, Sandra; Bryant, Adam; Larmer, Peter. The lower-limb tasks questionnaire: an assessment of validity, reliability, responsiveness, and minimal important differences. Arch Phys Med Rehabil. 2007;88(8):993-1001.

Oxford Knee Score (OKS)

Harris KKD, J.; Jones, L. D.; Beard, D. J.; Price, A. J. Extending the use of PROMs in the NHS-using the Oxford Knee Score in patients undergoing non-operative management for knee osteoarthritis: A validation study. BMJ Open. 2013;3(8).

Pain intensity numerical rating scale (NRS)

Salaffi FS, Andrea; Silvestri, Carlo Alberto; Ciapetti, Alessandro; Grassi, Walter. Minimal clinically important changes in chronic musculoskeletal pain intensity measured on a numerical rating scale. Eur J Pain. 2004;8(4):283-291.

Pain numerical rating scale (NRS)

Tubach FR, P.; Martin-Mola, E.; Awada, H.; Bellamy, N.; Bombardier, C.; Felson, D. T.; Hajjaj-Hassouni, N.; Hochberg, M.; Logeart, I.; Matucci-Cerinic, M.; van de Laar, M.; van der Heijde, D.; Dougados, M. Minimum clinically important improvement and patient acceptable symptom state in pain and function in rheumatoid arthritis, ankylosing spondylitis, chronic back pain, hand osteoarthritis, and hip and knee osteoarthritis: Results from a prospective multinational study. Arthritis Care Res (Hoboken). 2012;64(11):1699-1707.

Abbott JHS, John. Minimum Important Differences for the Patient-Specific Functional Scale, 4 Region-Specific Outcome Measures, and the Numeric Pain Rating Scale. J Orthop Sports Phys Ther. 2014;44(8):560-564.

Pain on movement during 48 hours before visit visual analogue scale (VAS)

Tubach FR, P.; Baron, G.; Falissard, B.; Logeart, I.; Bellamy, N.; Bombardier, C.; Felson, D.; Hochberg, M.; van der Heijde, D.; Dougados, M. Evaluation of clinically relevant changes in patient reported outcomes in knee and hip osteoarthritis: the minimal clinically important improvement. Ann Rheum Dis. 2005;64(1):29-33.

Pain on movement numerical rating scale (NRS)

Tubach FD, Maxime; Falissard, Bruno; Baron, Gabriel; Logeart, Isabelle; Ravaud, Philippe. Feeling good rather than feeling better matters more to patients. Arthritis Rheum. 2006;55(4):526-530.

Pain on movement visual analogue scale (VAS)

Tubach FD, Maxime; Falissard, Bruno; Baron, Gabriel; Logeart, Isabelle; Ravaud, Philippe. Feeling good rather than feeling better matters more to patients. Arthritis Rheum. 2006;55(4):526-530.

Patient-Rated Wrist Evaluation (PRWE)

Sorensen AAH, D.; Tan, W. H.; Ketchersid, J.; Calfee, R. P. Minimal clinically important differences of 3 patient-rated outcomes instruments. J Hand Surg [Am]. 2013;38(4):641-649.

Patient-Specific functional scale (PSFS)

Abbott JHS, John. Minimum Important Differences for the Patient-Specific Functional Scale, 4 Region-Specific Outcome Measures, and the Numeric Pain Rating Scale. J Orthop Sports Phys Ther. 2014;44(8):560-564.

Rheumatoid Arthritis Work Instability Scale (RA-WIS)

Roy J-SM, Joy C.; Amick, Benjamin C., 3rd; Shannon, Harry S.; McMurtry, Robert; Roth, James H.; Grewal, Ruby; Tang, Kenneth; Beaton, Dorcas. Validity and responsiveness of presenteeism scales in chronic work-related upper-extremity disorders. Phys Ther. 2011;91(2):254-266.

Short-Form Six-Dimension (SF-6D)

Walters SJB, John E. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. Qual Life Res. 2005;14(6):1523-1532.

Western Ontario and McMaster Universities Arthritis Index (WOMAC)

Tubach FR, P.; Baron, G.; Falissard, B.; Logeart, I.; Bellamy, N.; Bombardier, C.; Felson, D.; Hochberg, M.; van der Heijde, D.; Dougados, M. Evaluation of clinically relevant changes in patient reported outcomes in knee and hip osteoarthritis: the minimal clinically important improvement. Ann Rheum Dis. 2005;64(1):29-33.

Terwee CBR, Leo D.; Dekker, Joost; Bierma-Zeinstra, Sita M.; Peat, George; Jordan, Kelvin P.; Croft, Peter; de Vet, Henrica C. W. Mind the MIC: large variation among populations and methods. J Clin Epidemiol. 2010;63(5):524-534.

Angst FA, Andre; Michel, Beat A.; Stucki, Gerold. Minimal clinically important rehabilitation effects in patients with osteoarthritis of the lower extremities. J Rheumatol. 2002;29(1):131-138.

Ornetti PD, Maxime; Paternotte, Simon; Logeart, Isabelle; Gossec, Laure. Validation of a numerical rating scale to assess functional impairment in hip and knee osteoarthritis: comparison with the WOMAC function scale. Ann Rheum Dis. 2011;70(5):740-746.

Work Limitations Questionnaire (WLQ-25)

Roy J-SM, Joy C.; Amick, Benjamin C., 3rd; Shannon, Harry S.; McMurtry, Robert; Roth, James H.; Grewal, Ruby; Tang, Kenneth; Beaton, Dorcas. Validity and responsiveness of presenteeism scales in chronic work-related upper-extremity disorders. Phys Ther. 2011;91(2):254-266.

Work, Osteoarthritis or joint-Replacement Questionnaire (WORQ)

Kievit AJK, P. P.; Kievit, R. A.; Sierevelt, I. N.; Blankevoort, L.; Frings-Dresen, M. H. A reliable, valid and responsive questionnaire to score the impact of knee complaints on work following total knee arthroplasty: the WORQ. J Arthroplasty. 2014;29(6):1169-1175.e1162.

Other (Studies including patients with various conditions; Healthy individuals)

EuroQol-5D Utility Index (EQ-5D)
Walters SJB, John E. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. Qual Life Res. 2005;14(6):1523-1532.

Health related quality of life visual analogue scale (VAS)
Nichol MBE, Joshua D. Separating gains and losses in health when calculating the minimum important difference for mapped utility measures. Qual Life Res. 2008;17(6):955-961.

Health Utilities Index Mark II (HUI-II)
Samsa GE, D.; Rothman, M. L.; Williams, G. R.; Lipscomb, J.; Matchar, D. Determining clinically important differences in health status measures: a general approach with illustration to the Health Utilities Index Mark II. Pharmacoeconomics. 1999;15(2):141-155.

Health Utilities Index Mark III (HUI-III)
Wiebe SM, S.; Eliasziw, M.; Derry, P. A. Clinically important change in quality of life in epilepsy. J Neurol Neurosurg Psychiatry. 2002;73(2):116-120.

Measure Yourself Medical Outcome Profile (MYMOP) - Chinese version
Chung VCHW, V. C. W.; Lau, C. H.; Hui, H.; Lam, T. H.; Zhong, L. X.; Wong, S. Y. S.; Griffiths, S. M. Using Chinese Version of MYMOP in Chinese Medicine Evaluation: Validity, Responsiveness and Minimally Important Change. Health and Quality of Life Outcomes. 2010;8(111).

Quality of life in epilepsy inventory-31 (QOLIE-31)
Wiebe SM, S.; Eliasziw, M.; Derry, P. A. Clinically important change in quality of life in epilepsy. J Neurol Neurosurg Psychiatry. 2002;73(2):116-120.

Quality of life in epilepsy inventory-89 (QOLIE-89)
Wiebe SM, S.; Eliasziw, M.; Derry, P. A. Clinically important change in quality of life in epilepsy. J Neurol Neurosurg Psychiatry. 2002;73(2):116-120.

Short Form Health Survey 36-Item (SF-36)
Wiebe SM, S.; Eliasziw, M.; Derry, P. A. Clinically important change in quality of life in epilepsy. J Neurol Neurosurg Psychiatry. 2002;73(2):116-120.

Nichol MBE, Joshua D. Separating gains and losses in health when calculating the minimum important difference for mapped utility measures. Qual Life Res. 2008;17(6):955-961.

Rouquette AB, M.; Sebille, V.; Guillemin, F.; Cote, S. M.; Falissard, B.; Hardouin, J. B. The minimal clinically important difference determined using item response theory models: an attempt to solve the issue of the association with baseline score. J Clin Epidemiol. 2014;67(4):433-440.

Short-Form Six-Dimension (SF-6D)
Walters SJB, John E. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. Qual Life Res. 2005;14(6):1523-1532.

Walters SJB, John E. What is the relationship between the minimally important difference and health state utility values? The case of the SF-6D. Health Qual Life Outcomes. 2003;1:4.

Spanish society of contraception quality-of-life (SECQOL)
Perez-Campos ED, Jose Luis; de la Viuda, Esther; Gomez, Maria Angeles; Lertxundi, Roberto; Sanchez-Borrego, Rafael; Canals, Ignaci; Bermejo, Rafael; Arbat, Agnes; Badia, Xavier;

Perulero, Nuria; Lete, Luis Ignacio. Development and validation of the SEC-QOL questionnaire in women using contraceptive methods. Value Health. 2011;14(6):892-899.

Treatment Satisfaction With Medicines Questionnaire (SATMED-Q)

Rejas JR, Miguel A.; Pardo, Antonio; Soto, Javier. Minimally important difference of the Treatment Satisfaction with Medicines Questionnaire (SATMED-Q). BMC Med Res Methodol. 2011;11:142.

Plastic Surgery

Rhinoplasty Outcomes Evaluation (ROE) - Brazilian Portuguese version

Izu SCK, E. M.; Lopes, A. S.; Brandao, K. V.; Sousa, L. B.; Suguri, V. M.; Gregorio, L. C. Validation of the Rhinoplasty Outcomes Evaluation (ROE) questionnaire adapted to Brazilian Portuguese. Qual Life Res. 2014;23(3):953-958.

Psychiatry

Beck Depression Inventory-Second Edition (BDI-II) - Japanese version

Hiroe TK, M.; Yamamoto, I.; Nojima, S.; Kinoshita, Y.; Hashimoto, N.; Watanabe, N.; Maeda, T.; Furukawa, T. A. Gradations of clinical severity and sensitivity to change assessed with the Beck Depression Inventory-II in Japanese patients with depression. Psychiatry Res. 2005;135(3):229-235.

EuroQol-5D Utility Index (EQ-5D) - US weights

Le QAD, J. N.; Zoellner, L. A.; Feeny, N. C. Minimal clinically important differences for the EQ-5D and QWB-SA in Post-traumatic Stress Disorder (PTSD): results from a Doubly Randomized Preference Trial (DRPT). Health Qual Life Outcomes. 2013;11:59.

Health-Related Quality of Life for Eating Disorders questionnaire version-2 (HeRQoLEDv2)

Las Hayas CQ, Jose M.; Padierna, Jesus A.; Bilbao, Amaia; Munoz, Pedro; Francis Cook, E. Health-Related Quality of Life for Eating Disorders questionnaire version-2 was responsive 1-year after initial assessment. J Clin Epidemiol. 2007;60(8):825-833.

Insomnia Severity Index (ISI)

Morin CMB, Genevieve; Belanger, Lynda; Ivers, Hans. The Insomnia Severity Index: psychometric indicators to detect insomnia cases and evaluate treatment response. Sleep. 2011;34(5):601-608.

Medication Satisfaction Questionnaire (MSQ)

Vernon MKR, Dennis A.; Awad, A. George; Dirani, Riad; Panish, Jessica; Canuso, Carla M.; Grinspan, Augusto; Mannix, Sally; Kalali, Amir H. Psychometric evaluation of the Medication Satisfaction Questionnaire (MSQ) to assess satisfaction with antipsychotic medication among schizophrenia patients. Schizophr Res. 2010;118(1-3):271-278.

Quality of Well Being Self-Administered (QWB-SA)

Le QAD, J. N.; Zoellner, L. A.; Feeny, N. C. Minimal clinically important differences for the EQ-5D and QWB-SA in Post-traumatic Stress Disorder (PTSD): results from a Doubly Randomized Preference Trial (DRPT). Health Qual Life Outcomes. 2013;11:59.

Satisfactory Sexual Events (SSEs)

Symonds TS, Cathie; Sisson, Melanie; Soni, Paresh; Martin, Mona; Gunter, Lacey; Patrick, Donald L. Methods to determine the minimum important difference for a sexual event diary used by postmenopausal women with hypoactive sexual desire disorder. J Sex Med. 2007;4(5):1328-1335.

Respirology

Asthma Control Questionnaire (ACQ)

Juniper EFS, Klas; Mork, Ann-Christin; Stahl, Elisabeth. Measurement properties and interpretation of three shortened versions of the asthma control questionnaire. Respir Med. 2005;99(5):553-558.

Picado CB, C.; Perulero, N.; Sastre, J.; Olaguibel, J. M.; Lopez Vina, A.; Vega, J. M. Validation of the spanish version of the asthma control questionnaire. Clin Ther. 2008;30(10):1918-1931.

Asthma Quality of Life Questionnaire (AQLQ)

Juniper EFG, G. H.; Willan, A.; Griffith, L. E. Determining a minimal important change in a disease-specific Quality of Life Questionnaire. J Clin Epidemiol. 1994;47(1):81-87.

Mancuso CAP, Margaret G. E. Different methods to assess quality of life from multiple follow-ups in a longitudinal asthma study. J Clin Epidemiol. 2004;57(1):45-54.

Barber BLS, N. C.; Epstein, R. S. Impact of the global on patient perceivable change in an asthma specific QOL questionnaire. Qual Life Res. 1996;5(1):117-122.

Wyrwich KWM, Stacie M.; Kroenke, Kurt; Tierney, William M.; Babu, Ajit N.; Wolinsky, Fredric D. Interpreting quality-of-life data: methods for community consensus in asthma. Ann Allergy Asthma Immunol. 2006;96(6):826-833.

Chronic Respiratory Disease Questionnaire (CRQ)

Redelmeier DAG, G. H.; Goldstein, R. S. Assessing the minimal important difference in symptoms: a comparison of two techniques. J Clin Epidemiol. 1996;49(11):1215-1219.

Wyrwich KWM, Stacie M.; Kroenke, Kurt; Tierney, William M.; Babu, Ajit N.; Wolinsky, Fredric D. Measuring patient and clinician perspectives to evaluate change in health-related quality of life among patients with chronic obstructive pulmonary disease. J Gen Intern Med. 2007;22(2):161-170.

Turner DS, Holger J.; Griffith, Lauren E.; Beaton, Dorcas E.; Griffiths, Anne M.; Critch, Jeffrey N.; Guyatt, Gordon H. Using the entire cohort in the receiver operating characteristic analysis maximizes precision of the minimal important difference. J Clin Epidemiol. 2009;62(4):374-379.

Chronic Respiratory Disease Questionnaire (CRQ) / Chronic Heart Failure Questionnaire (CHQ)

Jaeschke RS, J.; Guyatt, G. H. Measurement of health status. Ascertaining the minimal clinically important difference. Control Clin Trials. 1989;10(4):407-415.

Clinical COPD Questionnaire (CCQ)

Kocks JWHT, M. G.; Uil, S. M.; van den Berg, J. W. K.; Stahl, E.; van der Molen, T. Health status measurement in COPD: the minimal clinically important difference of the clinical COPD questionnaire. Respir Res. 2006;7:62.

Kon SSD, D.; Mittal, M.; Nolan, C. M.; Clark, A. L.; Canavan, J. L.; Jones, S. E.; Polkey, M. I.; Man, W. D. The Clinical COPD Questionnaire: response to pulmonary rehabilitation and minimal clinically important difference. Thorax. 2014;69(9):793-798.

Canavan JLD, D.; Clark, A. L.; Jones, S. E.; Nolan, C. M.; Kon, S. S. C.; Man, W. D. C. Clinical COPD questionnaire in patients with chronic respiratory disease. Respirology. 2014;19(7):1006-1012.

COPD Assessment Test (CAT)

Kon SSC, J. L.; Jones, S. E.; Nolan, C. M.; Clark, A. L.; Dickson, M. J.; Haselden, B. M.; Polkey, M. I.; Man, W. D. Minimum clinically important difference for the COPD Assessment Test: a prospective analysis. The Lancet Respiratory Medicine. 2014;2(3):195-203.

Cough Quality of Life Questionnaire (CQLQ)

Fletcher KEF, Cynthia T.; Irwin, Richard S.; Corapi, Kristin M.; Norman, Geoffrey R. A prospective global measure, the Punum Ladder, provides more valid assessments of quality of life than a retrospective transition measure. J Clin Epidemiol. 2010;63(10):1123-1131.

Daytime Asthma Symptom Score

Santanello NCZ, J.; Seidenberg, B.; Reiss, T. F.; Barber, B. L. What are minimal important changes for asthma measures in a clinical trial? Eur Respir J. 1999;14(1):23-27.

Dyspnea visual analogue scale (VAS)

Karras DJS, M. E.; Terregino, C. A.; Lopez, B. L.; Griswold, S. K.; Arnold, G. K. Clinically meaningful changes in quantitative measures of asthma severity. Acad Emerg Med. 2000;7(4):327-334.

EuroQol-5D Utility Index (EQ-5D)

Walters SJB, John E. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. Qual Life Res. 2005;14(6):1523-1532.

Feeling thermometer (FT)

Schunemann HJG, Lauren; Jaeschke, Roman; Goldstein, Roger; Stubbing, David; Guyatt, Gordon H. Evaluation of the minimal important difference for the feeling thermometer and the St. George's Respiratory Questionnaire in patients with chronic airflow obstruction. J Clin Epidemiol. 2003;56(12):1170-1176.

Hospital Anxiety and Depression Scale (HADS)

Puhan MAF, M.; Buchi, S.; Schunemann, H. J. The minimal important difference of the hospital anxiety and depression scale in patients with chronic obstructive pulmonary disease. Health and Quality of Life Outcomes. 2008;6(46).

King's Brief ILD (K-BILD) questionnaire

Patel ASS, R. J.; Keir, G. J.; Bajwah, S.; Barker, R. D.; Maher, T. M.; Renzoni, E. A.; Wells, A. U.; Higginson, I. J.; Birring, S. S. The minimal important difference of the King's Brief Interstitial Lung Disease Questionnaire (K-BILD) and forced vital capacity in interstitial lung disease. Respir Med. 2013;107(9):1438-1443.

Quality of life for respiratory illness questionnaire (QoL-RIQ)

van Stel HFM, A. Rianne; Colland, Vivian T.; Everaerd, Walter. Interpretation of change and longitudinal validity of the quality of life for respiratory illness questionnaire (QoLRIQ) in inpatient pulmonary rehabilitation. Qual Life Res. 2003;12(2):133-145.

Quality of Life-Bronchiectasis (QOL-B V3.0)

Quittner ALOD, A. E.; Salathe, M. A.; Lewis, S. A.; Li, X.; Montgomery, A. B.; O'Riordan, T. G.; Barker, A. F. Quality of Life Questionnaire-Bronchiectasis: Final psychometric analyses and determination of minimal important difference scores. Thorax. 2015;70(1):12-20.

Short Form Health Survey 36-Item (SF-36)

Wyrwich KWM, Stacie M.; Kroenke, Kurt; Tierney, William M.; Babu, Ajit N.; Wolinsky, Fredric D. Measuring patient and clinician perspectives to evaluate change in health-related quality of life among patients with chronic obstructive pulmonary disease. J Gen Intern Med. 2007;22(2):161-170.

Short Form Health Survey 36-Item (SF-36) - version 2

Wyrwich KWM, Stacie M.; Kroenke, Kurt; Tierney, William M.; Babu, Ajit N.; Wolinsky, Fredric D. Interpreting quality-of-life data: methods for community consensus in asthma. Ann Allergy Asthma Immunol. 2006;96(6):826-833.

Short-Form Six-Dimension (SF-6D)

Walters SJB, John E. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. Qual Life Res. 2005;14(6):1523-1532.

Shortness of Breath with Daily Activities (SOBDA) Questionnaire

Watkins MLW, T. K.; Tabberer, M.; Brooks, J. M.; Donohue, J. F.; Anzueto, A.; Chen, W. H.; Crim, C. Shortness of breath with daily activities questionnaire: Validation and responder thresholds in patients with chronic obstructive pulmonary disease. BMJ Open. 2013;3(10).

St. George's Respiratory Questionnaire (SGRQ)

Schunemann HJG, Lauren; Jaeschke, Roman; Goldstein, Roger; Stubbing, David; Guyatt, Gordon H. Evaluation of the minimal important difference for the feeling thermometer and the St. George's Respiratory Questionnaire in patients with chronic airflow obstruction. J Clin Epidemiol. 2003;56(12):1170-1176.

St. George's Respiratory Questionnaire (SGRQ) - Mandarin-Chinese version

Xu WC, J. P.; Shapiro, S.; Lin, Y.; Yang, T.; Wang, C.; Bourbeau, J. Validation and clinical interpretation of the St George's Respiratory Questionnaire among COPD patients, China. Int J Tuberc Lung Dis. 2009;13(2):181-189.

University of California, San Diego Shortness of Breath Questionnaire (SOBQ)

Kupferberg DHK, Robert M.; Slymen, Donald J.; Ries, Andrew L. Minimal clinically important difference for the UCSD Shortness of Breath Questionnaire. J Cardiopulm Rehabil. 2005;25(6):370-377.

Visual Simplified Respiratory Questionnaire (VSRQ)

Perez TA, B.; Grosbois, J. M.; Bosch, V.; Guillemin, I.; Bravo, M. L.; Brun, M.; Tonnel, A. B.; Group, Tiphon Study. Validity, reliability, and responsiveness of a new short Visual Simplified Respiratory Questionnaire (VSRQ) for health-related quality of life assessment in chronic obstructive pulmonary disease. Int J Chron Obstruct Pulmon Dis. 2009;4:9-18.

Respirology, Rheumatology

Transition Dyspnoea Index (TDI)

Khanna DT, Chi-Hong; Furst, Daniel E.; Clements, Philip J.; Elashoff, Robert; Roth, Michael; Elashoff, David; Tashkin, Donald P.; for Scleroderma Lung Study, Investigators. Minimally important differences in the Mahler's Transition Dyspnoea Index in a large randomized controlled trial--results from the Scleroderma Lung Study. Rheumatology (Oxford). 2009;48(12):1537-1540.

Rheumatology

Bath Ankylosing Spondylitis Disease Activity Index (BASDAI)

Pavy SB, Sinead; Calin, Andrei. Establishment of the minimum clinically important difference for the bath ankylosing spondylitis indices: a prospective study. J Rheumatol. 2005;32(1):80-85.

Bath Ankylosing Spondylitis Functional Index (BASFI)

Pavy SB, Sinead; Calin, Andrei. Establishment of the minimum clinically important difference for the bath ankylosing spondylitis indices: a prospective study. J Rheumatol. 2005;32(1):80-85.

Bath Ankylosing Spondylitis Global Score (BAS-G)

Pavy SB, Sinead; Calin, Andrei. Establishment of the minimum clinically important difference for the bath ankylosing spondylitis indices: a prospective study. J Rheumatol. 2005;32(1):80-85.

Chalder Fatigue Scale (CFS)

Pouchot JK, Raheem B.; Brant, Rollin; Lacaille, Diane; Lehman, Allen J.; Ensworth, Stephanie; Kopec, Jacek; Esdaile, John M.; Liang, Matthew H. Determination of the minimal clinically important difference for seven fatigue measures in rheumatoid arthritis. J Clin Epidemiol. 2008;61(7):705-713.

Goligher ECP, Jacques; Brant, Rollin; Kherani, Raheem B.; Avina-Zubieta, J. Antonio; Lacaille, Diane; Lehman, Allen J.; Ensworth, Stephanie; Kopec, Jacek; Esdaile, John M.; Liang, Matthew H. Minimal clinically important difference for 7 measures of fatigue in patients with systemic lupus erythematosus. J Rheumatol. 2008;35(4):635-642.

Disease activity numerical rating scale (NRS)

Tubach FR, P.; Martin-Mola, E.; Awada, H.; Bellamy, N.; Bombardier, C.; Felson, D. T.; Hajjaj-Hassouni, N.; Hochberg, M.; Logeart, I.; Matucci-Cerinic, M.; van de Laar, M.; van der Heijde, D.; Dougados, M. Minimum clinically important improvement and patient acceptable symptom state in pain and function in rheumatoid arthritis, ankylosing spondylitis, chronic back pain, hand osteoarthritis, and hip and knee osteoarthritis: Results from a prospective multinational study. Arthritis Care Res (Hoboken). 2012;64(11):1699-1707.

EuroQol-5D Utility Index (EQ-5D)

Walters SJB, John E. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. Qual Life Res. 2005;14(6):1523-1532.

Kvamme MKK, Ivar Sonbo; Lie, Elisabeth; Kvien, Tore Kristian. Identification of cutpoints for acceptable health status and important improvement in patient-reported outcomes, in rheumatoid arthritis, psoriatic arthritis, and ankylosing spondylitis. J Rheumatol. 2010;37(1):26-31.

Kwakkenbos LF, J.; Vonk, M. C.; Becker, E. S.; Jeurissen, M.; van den Hoogen, F. H.; van den Ende, C. H. A comparison of the measurement properties and estimation of minimal important differences of the EQ-5D and SF-6D utility measures in patients with systemic sclerosis. Clin Exp Rheumatol. 2013;31(2 Suppl 76):50-56.

Fatigue Assessment Scale (FAS)

de Kleijn WPEDV, Jolanda; Wijnen, Petal A. H. M.; Drent, Marjolein. Minimal (clinically) important differences for the Fatigue Assessment Scale in sarcoidosis. Respir Med. 2011;105(9):1388-1395.

Fatigue severity scale (FSS)

Pouchot JK, Raheem B.; Brant, Rollin; Lacaille, Diane; Lehman, Allen J.; Ensworth, Stephanie; Kopec, Jacek; Esdaile, John M.; Liang, Matthew H. Determination of the minimal clinically important difference for seven fatigue measures in rheumatoid arthritis. J Clin Epidemiol. 2008;61(7):705-713.

Goligher ECP, Jacques; Brant, Rollin; Kherani, Raheem B.; Avina-Zubieta, J. Antonio; Lacaille, Diane; Lehman, Allen J.; Ensworth, Stephanie; Kopec, Jacek; Esdaile, John M.; Liang, Matthew H. Minimal clinically important difference for 7 measures of fatigue in patients with systemic lupus erythematosus. J Rheumatol. 2008;35(4):635-642.

Fatigue visual analogue scale (VAS)

Wells GL, Tracy; Maxwell, Lara; MacLean, Ross; Tugwell, Peter. Determining the minimal clinically important differences in activity, fatigue, and sleep quality in patients with rheumatoid arthritis. J Rheumatol. 2007;34(2):280-289.

Wheaton LP, Janet. The minimally important difference for patient-reported outcomes in spondyloarthropathies including pain, fatigue, sleep, and Health Assessment Questionnaire. J Rheumatol. 2010;37(4):816-822.

Khanna DP, Janet E.; Khanna, Puja P.; Maloney, Michelle; Samedi, Nooshin; Norrie, Debbie; Ouimet, Gillian; Hays, Ron D. The minimally important difference for the fatigue visual analog scale in patients with rheumatoid arthritis followed in an academic clinical practice. J Rheumatol. 2008;35(12):2339-2343.

Sekhon SP, Janet; Canadian Scleroderma Research, Group; Baron, Murray. The minimally important difference in clinical practice for patient-centered outcomes including health assessment questionnaire, fatigue, pain, sleep, global visual analog scale, and SF-36 in scleroderma. J Rheumatol. 2010;37(3):591-598.

George AP, J. E. The minimally important difference (MID) for patient-reported outcomes including pain, fatigue, sleep and the health assessment questionnaire disability index (HAQ-DI) in primary Sjogren's syndrome. Clin Exp Rheumatol. 2011;29(2):248-253.

Colangelo KJP, Janet E.; Peschken, Christine. The minimally important difference for patient reported outcomes in systemic lupus erythematosus including the HAQ-DI, pain, fatigue, and SF-36. J Rheumatol. 2009;36(10):2231-2237.

Kwok TP, Janet E. Minimally important difference for patient-reported outcomes in psoriatic arthritis: Health Assessment Questionnaire and pain, fatigue, and global visual analog scales. J Rheumatol. 2010;37(5):1024-1028.

Fibromyalgia Impact Questionnaire (FIQ)

Bennett RMB, Andrew G.; Cappelleri, Joseph C.; Zlateva, Gergana; Sadosky, Alesia B. Minimal clinically important difference in the fibromyalgia impact questionnaire. J Rheumatol. 2009;36(6):1304-1311.

Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT-F)

Pouchot JK, Raheem B.; Brant, Rollin; Lacaille, Diane; Lehman, Allen J.; Ensworth, Stephanie; Kopec, Jacek; Esdaile, John M.; Liang, Matthew H. Determination of the minimal clinically important difference for seven fatigue measures in rheumatoid arthritis. J Clin Epidemiol. 2008;61(7):705-713.

Strand VK, M.; Gnanasakthy, A.; Mallya, U.; Mpofu, S. Secukinumab treatment in rheumatoid arthritis is associated with incremental benefit in the clinical outcomes and HRQoL improvements that exceed minimally important thresholds. Health Qual Life Outcomes. 2014;12:31.

Goligher ECP, Jacques; Brant, Rollin; Kherani, Raheem B.; Avina-Zubieta, J. Antonio; Lacaille, Diane; Lehman, Allen J.; Ensworth, Stephanie; Kopec, Jacek; Esdaile, John M.; Liang, Matthew H. Minimal clinically important difference for 7 measures of fatigue in patients with systemic lupus erythematosus. J Rheumatol. 2008;35(4):635-642.

Lai J-SB, Jennifer L.; Ogale, Sarika; Brunetta, Paul; Cella, David. Validation of the functional assessment of chronic illness therapy-fatigue scale in patients with moderately to severely active systemic lupus erythematosus, participating in a clinical trial. J Rheumatol. 2011;38(4):672-679.

Functional disability numerical rating scale (NRS)

Tubach FR, P.; Martin-Mola, E.; Awada, H.; Bellamy, N.; Bombardier, C.; Felson, D. T.; Hajjaj-Hassouni, N.; Hochberg, M.; Logeart, I.; Matucci-Cerinic, M.; van de Laar, M.; van der Heijde, D.; Dougados, M. Minimum clinically important improvement and patient acceptable symptom state in pain and function in rheumatoid arthritis, ankylosing spondylitis, chronic back pain, hand osteoarthritis, and hip and knee osteoarthritis: Results from a prospective multinational study. Arthritis Care Res (Hoboken). 2012;64(11):1699-1707.

Global assessment of fatigue numerical rating scale (NRS)

Pouchot JK, Raheem B.; Brant, Rollin; Lacaille, Diane; Lehman, Allen J.; Ensworth, Stephanie; Kopec, Jacek; Esdaile, John M.; Liang, Matthew H. Determination of the minimal clinically important difference for seven fatigue measures in rheumatoid arthritis. J Clin Epidemiol. 2008;61(7):705-713.

Goligher ECP, Jacques; Brant, Rollin; Kherani, Raheem B.; Avina-Zubieta, J. Antonio; Lacaille, Diane; Lehman, Allen J.; Ensworth, Stephanie; Kopec, Jacek; Esdaile, John M.; Liang, Matthew H. Minimal clinically important difference for 7 measures of fatigue in patients with systemic lupus erythematosus. J Rheumatol. 2008;35(4):635-642.

Global health status visual analogue scale (VAS)

Kvamme MKK, Ivar Sonbo; Lie, Elisabeth; Kvien, Tore Kristian. Identification of cutpoints for acceptable health status and important improvement in patient-reported outcomes, in rheumatoid arthritis, psoriatic arthritis, and ankylosing spondylitis. J Rheumatol. 2010;37(1):26-31.

Wheaton LP, Janet. The minimally important difference for patient-reported outcomes in spondyloarthropathies including pain, fatigue, sleep, and Health Assessment Questionnaire. J Rheumatol. 2010;37(4):816-822.

Sekhon SP, Janet; Canadian Scleroderma Research, Group; Baron, Murray. The minimally important difference in clinical practice for patient-centered outcomes including health assessment questionnaire, fatigue, pain, sleep, global visual analog scale, and SF-36 in scleroderma. J Rheumatol. 2010;37(3):591-598.

Ward MMG, L. C.; Alba, M. 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. J Clin Epidemiol. 2014;67(6):689-696.

George AP, J. E. The minimally important difference (MID) for patient-reported outcomes including pain, fatigue, sleep and the health assessment questionnaire disability index (HAQ-DI) in primary Sjogren's syndrome. Clin Exp Rheumatol. 2011;29(2):248-253.

Kwok TP, Janet E. Minimally important difference for patient-reported outcomes in psoriatic arthritis: Health Assessment Questionnaire and pain, fatigue, and global visual analog scales. J Rheumatol. 2010;37(5):1024-1028.

Gout Assessment Questionnaire (GAQ)

Colwell HHH, B. J.; Pasta, D. J.; Palo, W. A.; Mathias, S. D.; Joseph-Ridge, N. Gout Assessment Questionnaire: Initial results of reliability, validity and responsiveness. Int J Clin Pract. 2006;60(10):1210-1217.

Gout Impact Scale (GIS)

Khanna DS, Andrew J.; Khanna, Puja P.; Shieh, Marian M.; Kavanaugh, Arthur F.; Terkeltaub, Robert A.; Lee, Susan J.; Singh, Jasvinder A.; Hirsch, Jan D. Minimally important differences of the gout impact scale in a randomized controlled trial. Rheumatology (Oxford). 2011;50(7):1331-1336.

Health Assessment Questionnaire (HAQ) Disability Index

Kosinski MZ, S. Z.; Dedhiya, S.; Osterhaus, J. T.; Ware, J. E., 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(7):1478-1487.

Wheaton LP, Janet. The minimally important difference for patient-reported outcomes in spondyloarthropathies including pain, fatigue, sleep, and Health Assessment Questionnaire. J Rheumatol. 2010;37(4):816-822.

Pope JEK, Dinesh; Norrie, Deborah; Ouimet, Janine M. 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(2):254-259.

Sekhon SP, Janet; Canadian Scleroderma Research, Group; Baron, Murray. The minimally important difference in clinical practice for patient-centered outcomes including health assessment questionnaire, fatigue, pain, sleep, global visual analog scale, and SF-36 in scleroderma. J Rheumatol. 2010;37(3):591-598.

Strand VK, M.; Gnanasakthy, A.; Mallya, U.; Mpofu, S. Secukinumab treatment in rheumatoid arthritis is associated with incremental benefit in the clinical outcomes and HRQoL improvements that exceed minimally important thresholds. Health Qual Life Outcomes. 2014;12:31.

George AP, J. E. The minimally important difference (MID) for patient-reported outcomes including pain, fatigue, sleep and the health assessment questionnaire disability index (HAQ-DI) in primary Sjogren's syndrome. Clin Exp Rheumatol. 2011;29(2):248-253.

Colangelo KJP, Janet E.; Peschken, Christine. The minimally important difference for patient reported outcomes in systemic lupus erythematosus including the HAQ-DI, pain, fatigue, and SF-36. J Rheumatol. 2009;36(10):2231-2237.

Kwok TP, Janet E. Minimally important difference for patient-reported outcomes in psoriatic arthritis: Health Assessment Questionnaire and pain, fatigue, and global visual analog scales. J Rheumatol. 2010;37(5):1024-1028.

Mease PJW, J. Michael; Bitman, Bojena; Wang, Brian C.; Globe, Denise R.; Singh, Amitabh. Minimally important difference of Health Assessment Questionnaire in psoriatic arthritis: relating thresholds of improvement in functional ability to patient-rated importance and satisfaction. J Rheumatol. 2011;38(11):2461-2465.

Leung YYZ, T. Y.; Tam, L. S.; Kun, E. W. L.; Li, E. K. M. Minimal important difference and responsiveness to change of the SF-36 in patients with psoriatic arthritis receiving tumor necrosis factor-alpha blockers. J Rheumatol. 2011;38(9):2077-2079.

Joint tenderness 4-point likert scale

Taylor WJR, D.; Dalbeth, N.; Schumacher, H. R.; Edwards, N. L.; Simon, L. S.; John, M. R.; Essex, M. N.; Watson, D. J.; Evans, R.; Rome, K.; Singh, J. A. Application of the OMERACT filter to measures of core outcome domains in recent clinical studies of acute gout. J Rheumatol. 2014;41(3):574-580.

Medical Outcomes Study Sleep Scale (MOS Sleep)

Wells GL, Tracy; Maxwell, Lara; MacLean, Ross; Tugwell, Peter. Determining the minimal clinically important differences in activity, fatigue, and sleep quality in patients with rheumatoid arthritis. J Rheumatol. 2007;34(2):280-289.

Modified Health Assessment Questionnaire (MHAQ) Disability Index

Kvamme MKK, Ivar Sonbo; Lie, Elisabeth; Kvien, Tore Kristian. Identification of cutpoints for acceptable health status and important improvement in patient-reported outcomes, in rheumatoid arthritis, psoriatic arthritis, and ankylosing spondylitis. J Rheumatol. 2010;37(1):26-31.

Multidimensional Assessment of Fatigue (MAF)

Pouchot JK, Raheem B.; Brant, Rollin; Lacaille, Diane; Lehman, Allen J.; Ensworth, Stephanie; Kopec, Jacek; Esdaile, John M.; Liang, Matthew H. Determination of the minimal clinically important difference for seven fatigue measures in rheumatoid arthritis. J Clin Epidemiol. 2008;61(7):705-713.

Goligher ECP, Jacques; Brant, Rollin; Kherani, Raheem B.; Avina-Zubieta, J. Antonio; Lacaille, Diane; Lehman, Allen J.; Ensworth, Stephanie; Kopec, Jacek; Esdaile, John M.; Liang, Matthew H. Minimal clinically important difference for 7 measures of fatigue in patients with systemic lupus erythematosus. J Rheumatol. 2008;35(4):635-642.

Multidimensional Fatigue Inventory (MFI)

Pouchot JK, Raheem B.; Brant, Rollin; Lacaille, Diane; Lehman, Allen J.; Ensworth, Stephanie; Kopec, Jacek; Esdaile, John M.; Liang, Matthew H. Determination of the minimal clinically important difference for seven fatigue measures in rheumatoid arthritis. J Clin Epidemiol. 2008;61(7):705-713.

Goligher ECP, Jacques; Brant, Rollin; Kherani, Raheem B.; Avina-Zubieta, J. Antonio; Lacaille, Diane; Lehman, Allen J.; Ensworth, Stephanie; Kopec, Jacek; Esdaile, John M.; Liang, Matthew H. Minimal clinically important difference for 7 measures of fatigue in patients with systemic lupus erythematosus. J Rheumatol. 2008;35(4):635-642.

Pain 5-point likert scale

Taylor WJR, D.; Dalbeth, N.; Schumacher, H. R.; Edwards, N. L.; Simon, L. S.; John, M. R.; Essex, M. N.; Watson, D. J.; Evans, R.; Rome, K.; Singh, J. A. Application of the OMERACT filter to measures of core outcome domains in recent clinical studies of acute gout. J Rheumatol. 2014;41(3):574-580.

Pain numerical rating scale (NRS)

Tubach FR, P.; Martin-Mola, E.; Awada, H.; Bellamy, N.; Bombardier, C.; Felson, D. T.; Hajjaj-Hassouni, N.; Hochberg, M.; Logeart, I.; Matucci-Cerinic, M.; van de Laar, M.; van der Heijde, D.; Dougados, M. Minimum clinically important improvement and patient acceptable symptom state in pain and function in rheumatoid arthritis, ankylosing spondylitis, chronic back pain, hand osteoarthritis, and hip and knee osteoarthritis: Results from a prospective multinational study. Arthritis Care Res (Hoboken). 2012;64(11):1699-1707.

Pain visual analogue scale (VAS)

Wolfe FM, Kaleb. Assessment of pain in rheumatoid arthritis: minimal clinically significant difference, predictors, and the effect of anti-tumor necrosis factor therapy. J Rheumatol. 2007;34(8):1674-1683.

Kvamme MKK, Ivar Sonbo; Lie, Elisabeth; Kvien, Tore Kristian. Identification of cutpoints for acceptable health status and important improvement in patient-reported outcomes, in rheumatoid arthritis, psoriatic arthritis, and ankylosing spondylitis. J Rheumatol. 2010;37(1):26-31.

Wheaton LP, Janet. The minimally important difference for patient-reported outcomes in spondyloarthropathies including pain, fatigue, sleep, and Health Assessment Questionnaire. J Rheumatol. 2010;37(4):816-822.

Pope JEK, Dinesh; Norrie, Deborah; Ouimet, Janine M. 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(2):254-259.

Sekhon SP, Janet; Canadian Scleroderma Research, Group; Baron, Murray. The minimally important difference in clinical practice for patient-centered outcomes including health assessment questionnaire, fatigue, pain, sleep, global visual analog scale, and SF-36 in scleroderma. J Rheumatol. 2010;37(3):591-598.

George AP, J. E. The minimally important difference (MID) for patient-reported outcomes including pain, fatigue, sleep and the health assessment questionnaire disability index (HAQ-DI) in primary Sjogren's syndrome. Clin Exp Rheumatol. 2011;29(2):248-253.

Colangelo KJP, Janet E.; Peschken, Christine. The minimally important difference for patient reported outcomes in systemic lupus erythematosus including the HAQ-DI, pain, fatigue, and SF-36. J Rheumatol. 2009;36(10):2231-2237.

Kwok TP, Janet E. Minimally important difference for patient-reported outcomes in psoriatic arthritis: Health Assessment Questionnaire and pain, fatigue, and global visual analog scales. J Rheumatol. 2010;37(5):1024-1028.

Leung YYZ, T. Y.; Tam, L. S.; Kun, E. W. L.; Li, E. K. M. Minimal important difference and responsiveness to change of the SF-36 in patients with psoriatic arthritis receiving tumor necrosis factor-alpha blockers. J Rheumatol. 2011;38(9):2077-2079.

Patient's global health assessment (PGA)

Leung YYZ, T. Y.; Tam, L. S.; Kun, E. W. L.; Li, E. K. M. Minimal important difference and responsiveness to change of the SF-36 in patients with psoriatic arthritis receiving tumor necrosis factor-alpha blockers. J Rheumatol. 2011;38(9):2077-2079.

PROMIS 20-item Physical Functioning Short Form (PROMIS PF-20)

Hays RDS, K. L.; Fries, J. F.; Krishnan, E. Responsiveness and minimally important difference for the Patient-Reported Outcomes Measurement Information System (PROMIS) 20-item physical functioning short form in a prospective observational study of rheumatoid arthritis. Ann Rheum Dis. 2015;74(1):104-107.

Raynaud's Condition Score (RCS) - Visual analogoue scale version

Khanna PPM, Paul; Gregory, Jeff; Khanna, Dinesh. The minimally important difference and patient acceptable symptom state for the Raynaud's condition score in patients with Raynaud's phenomenon in a large randomised controlled clinical trial. Ann Rheum Dis. 2010;69(3):588-591.

Revised Cedars-Sinai Health-Related Quality of Life for Rheumatoid Arthritis Instrument (CSHQ-RA)

Chiou CFS, C. D.; Cornelio, I.; Lubeck, D. P.; Paulus, H. E.; Dylan, M.; Chang, C. H.; Weisman, M. H. Development and validation of the revised Cedars-Sinai health-related quality of life for rheumatoid arthritis instrument. Arthritis Rheum. 2006;55(6):856-863.

Rheumatoid Arthritis Impact of Disease (RAID)

Dougados MB, Y.; Logeart, I.; van der Heijde, D.; Gossec, L.; Kvien, T. Defining cut-off values for disease activity states and improvement scores for patient-reported outcomes: The example of the Rheumatoid Arthritis Impact of Disease (RAID). Arthritis Research and Therapy. 2012;14(3).

Short Form Health Survey 36-Item (SF-36)

Pouchot JK, Raheem B.; Brant, Rollin; Lacaille, Diane; Lehman, Allen J.; Ensworth, Stephanie; Kopec, Jacek; Esdaile, John M.; Liang, Matthew H. Determination of the minimal clinically important difference for seven fatigue measures in rheumatoid arthritis. J Clin Epidemiol. 2008;61(7):705-713.

Sekhon SP, Janet; Canadian Scleroderma Research, Group; Baron, Murray. The minimally important difference in clinical practice for patient-centered outcomes including health assessment questionnaire, fatigue, pain, sleep, global visual analog scale, and SF-36 in scleroderma. J Rheumatol. 2010;37(3):591-598.

Goligher ECP, Jacques; Brant, Rollin; Kherani, Raheem B.; Avina-Zubieta, J. Antonio; Lacaille, Diane; Lehman, Allen J.; Ensworth, Stephanie; Kopec, Jacek; Esdaile, John M.; Liang, Matthew H. Minimal clinically important difference for 7 measures of fatigue in patients with systemic lupus erythematosus. J Rheumatol. 2008;35(4):635-642.

Colangelo KJP, Janet E.; Peschken, Christine. The minimally important difference for patient reported outcomes in systemic lupus erythematosus including the HAQ-DI, pain, fatigue, and SF-36. J Rheumatol. 2009;36(10):2231-2237.

Leung YYZ, T. Y.; Tam, L. S.; Kun, E. W. L.; Li, E. K. M. Minimal important difference and responsiveness to change of the SF-36 in patients with psoriatic arthritis receiving tumor necrosis factor-alpha blockers. J Rheumatol. 2011;38(9):2077-2079.

Short Form Health Survey 36-Item (SF-36) - version 2

Strand VK, M.; Gnanasakthy, A.; Mallya, U.; Mpofu, S. Secukinumab treatment in rheumatoid arthritis is associated with incremental benefit in the clinical outcomes and HRQoL improvements that exceed minimally important thresholds. Health Qual Life Outcomes. 2014;12:31.

Short-Form Six-Dimension (SF-6D)

Walters SJB, John E. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. Qual Life Res. 2005;14(6):1523-1532.

Khanna DF, Daniel E.; Wong, Weng Kee; Tsevat, Joel; Clements, Philip J.; Park, Grace S.; Postlethwaite, Arnold E.; Ahmed, Mansoor; Ginsburg, Shaari; Hays, Ron D.; Scleroderma Collagen Type 1 Study, Group. Reliability, validity, and minimally important differences of the SF-6D in systemic sclerosis. Qual Life Res. 2007;16(6):1083-1092.

Kvamme MKK, Ivar Sonbo; Lie, Elisabeth; Kvien, Tore Kristian. Identification of cutpoints for acceptable health status and important improvement in patient-reported outcomes, in rheumatoid arthritis, psoriatic arthritis, and ankylosing spondylitis. J Rheumatol. 2010;37(1):26-31.

Kwakkenbos LF, J.; Vonk, M. C.; Becker, E. S.; Jeurissen, M.; van den Hoogen, F. H.; van den Ende, C. H. A comparison of the measurement properties and estimation of minimal important differences of the EQ-5D and SF-6D utility measures in patients with systemic sclerosis. Clin Exp Rheumatol. 2013;31(2 Suppl 76):50-56.

Sleep visual analogue scale (VAS)

Wheaton LP, Janet. The minimally important difference for patient-reported outcomes in spondyloarthropathies including pain, fatigue, sleep, and Health Assessment Questionnaire. J Rheumatol. 2010;37(4):816-822.

Sekhon SP, Janet; Canadian Scleroderma Research, Group; Baron, Murray. The minimally important difference in clinical practice for patient-centered outcomes including health assessment questionnaire, fatigue, pain, sleep, global visual analog scale, and SF-36 in scleroderma. J Rheumatol. 2010;37(3):591-598.

George AP, J. E. The minimally important difference (MID) for patient-reported outcomes including pain, fatigue, sleep and the health assessment questionnaire disability index (HAQ-DI) in primary Sjogren's syndrome. Clin Exp Rheumatol. 2011;29(2):248-253.

Colangelo KJP, Janet E.; Peschken, Christine. The minimally important difference for patient reported outcomes in systemic lupus erythematosus including the HAQ-DI, pain, fatigue, and SF-36. J Rheumatol. 2009;36(10):2231-2237.

Kwok TP, Janet E. Minimally important difference for patient-reported outcomes in psoriatic arthritis: Health Assessment Questionnaire and pain, fatigue, and global visual analog scales. J Rheumatol. 2010;37(5):1024-1028.

Thoracic Surgery

University of California, San Diego Shortness of Breath Questionnaire (SOBQ)

Horita NM, N.; Morita, S.; Kojima, R.; Kimura, N.; Kaneko, T.; Ishigatsubo, Y. Small, moderate, and large changes, and the minimum clinically important difference in the University of California, San Diego Shortness of Breath Questionnaire. Copd. 2014;11(1):26-32.

Urology

Erection Quality Scale (EQS)

Rosen RCW, John; Mollen, Martin D.; Gondek, Kathleen; McLeod, Lori D.; Fisher, William A. Responsiveness and minimum important differences for the erection quality scale. J Urol. 2007;178(5):2076-2081.

Incontinence Impact Questionnaire Short Form (IIQ-7) adjusted

Utomo EK, I. J.; Wildhagen, M. F.; Steensma, A. B.; Bangma, C. H.; Blok, B. F. M. Validation of the urogenital distress inventory (UDI-6) and incontinence impact questionnaire (IIQ-7) in a Dutch population. Neurourology and Urodynamics. 2014;34(1):24-31.

Incontinence Impact Questionnaire Short Form (IIQ-7)

Utomo EK, I. J.; Wildhagen, M. F.; Steensma, A. B.; Bangma, C. H.; Blok, B. F. M. Validation of the urogenital distress inventory (UDI-6) and incontinence impact questionnaire (IIQ-7) in a Dutch population. Neurourology and Urodynamics. 2014;34(1):24-31.

Incontinence Impact Questionnaire-8 (IIQ-8)

Utomo EK, I. J.; Wildhagen, M. F.; Steensma, A. B.; Bangma, C. H.; Blok, B. F. M. Validation of the urogenital distress inventory (UDI-6) and incontinence impact questionnaire (IIQ-7) in a Dutch population. Neurourology and Urodynamics. 2014;34(1):24-31.

Interstitial Cystitis Symptom Index (ICSI)

Lubeck DPW, K.; Sant, G. R.; Alvarez-Horine, S.; Lai, C. Psychometric validation of the O'leary-Sant interstitial cystitis symptom index in a clinical trial of pentosan polysulfate sodium. Urology. 2001;57(6 Suppl 1):62-66.

King's Health Questionnaire (KHQ)

Kelleher CJP, A. M.; Reese, P. R.; Burgess, S. M.; Brodish, P. H. How much is enough and who says so? The case of the King's Health Questionnaire and overactive bladder. BJOG: An International Journal of Obstetrics and Gynaecology. 2004;111(6):605-612.

Michigan Incontinence Symptom Index (M-ISI)

Suskind AMD, R. L.; Morgan, D. M.; Delancey, J. O. L.; McGuire, E. J.; Wei, J. T. The Michigan incontinence symptom index (M-ISI): A clinical measure for type, severity, and bother related to urinary incontinence. Neurourology and Urodynamics. 2014;33(7):1128-1134.

Overactive Bladder Symptom And Health-Related Quality Of Life Questionnaire (OAB-Q)

Coyne KSM, Louis S.; Thompson, Christine L.; Kopp, Zoe S.; Khullar, Vikram. Determining the importance of change in the overactive bladder questionnaire. J Urol. 2006;176(2):627-632; discussion 632.

Overactive Bladder Symptom Score (OABSS)

Gotoh MH, Yukio; Yokoyama, Osamu; Nishizawa, Osamu. Responsiveness and minimal clinically important change in overactive bladder symptom score. Urology. 2011;78(4):768-773.

Sexual Experience Questionnaire (SEX-Q)

Mulhall JPK, Rosie; Kirby, Michael; Hvidsten, Kyle; Symonds, Tara; Bushmakin, Andrew G.; Cappelleri, Joseph C. Evaluating the sexual experience in men: validation of the sexual experience questionnaire. J Sex Med. 2008;5(2):365-376.

Urogenital Distress Inventory-Short Form (UDI-6)

Utomo EK, I. J.; Wildhagen, M. F.; Steensma, A. B.; Bangma, C. H.; Blok, B. F. M. Validation of the urogenital distress inventory (UDI-6) and incontinence impact questionnaire (IIQ-7) in a Dutch population. Neurourology and Urodynamics. 2014;34(1):24-31.

Vascular Surgery

Aberdeen Varicose Veins Questionnaire (AVVQ)

Browne JPvdM, Jan H.; Lewsey, James D.; Lamping, Donna L.; Black, Nick. Mathematical coupling may account for the association between baseline severity and minimally important difference values. J Clin Epidemiol. 2010;63(8):865-874.

EuroQol-5D Utility Index (EQ-5D)

Browne JPvdM, Jan H.; Lewsey, James D.; Lamping, Donna L.; Black, Nick. Mathematical coupling may account for the association between baseline severity and minimally important difference values. J Clin Epidemiol. 2010;63(8):865-874.

Specific Quality of Life & Outcome Response - Venous (SQOR-V)

Lurie FK, Robert L. In prospective study using Specific Quality of Life & Outcomes Response-Venous (SQOR-V) questionnaire the recall bias had the same magnitude as the minimally important difference. Qual Life Res. 2011;20(10):1589-1593.

Vascular Quality of Life questionnaire (VascuQol)

Frans FAN, P. T.; Met, R.; Bipat, S.; Legemate, D. A.; Reekers, J. A.; Koelemay, M. J. Statistical or clinical improvement? Determining the minimally important difference for the vascular quality of life questionnaire in patients with critical limb ischemia. Eur J Vasc Endovasc Surg. 2014;47(2):180-186.

Chapter 4: Mind the Methods of Determining Minimal Important Differences – Three critical issues to consider

Tahira Devji, Alonso Carrasco-Labra, Gordon H Guyatt

Submitted to: BMJ [Dec 2018]

ABSTRACT

Patient reported outcome measures (PROMs) are increasingly used in randomized clinical trials, meta-analyses and clinical practice guidelines. However, the use of PROMs involves challenges, the most important of which is deciding if a particular treatment effect is trivial, small but important, moderate or large. Of the strategies to facilitate interpretation of PROMs, the minimal important difference (MID), which provides a measure of the smallest change in an instrument score that patients consider important, has proved the most useful. Over the past three decades, researchers have developed various methods to estimate MIDs. However, the studies using these methods are not planned, conducted or reported equally well. Nevertheless, the field is expanding and new studies empirically estimating MIDs are being published very frequently. Our research group has developed resources for facilitating the identification and appraisal of anchor-based MIDs: a comprehensive inventory of anchor-based MIDs and an instrument to evaluate the credibility of these estimates. In this article, we elucidate three critical issues with current MID literature, informed by our prior work, and have identified opportunities for important advancements in the field.

To judge patients' response to therapy and inform clinical decision-making, clinicians rely, in daily practice on patients' self-assessment of change in health status. A typical question clinicians ask their patients is "Since last week when we started the new treatment, are you feeling better or worse – and if so, to what extent?". Such "transition questions" are single items that are short, simple to administer, easy to interpret and are therefore very appealing to clinicians, patients and other health-care stakeholders[1].

Although transition questions represent intuitively valuable patient-reported outcome measures (PROMs), classic measurement theory holds that such single item measures have important limitations relative to multi-item measures: they are less stable, reliable and precise[2]. Well-constructed multi-item instruments are also more sensitive to changes in patients' health status over time (i.e. responsive), and able to measure multidimensional phenomena and provide information regarding individual domains[2-5]. Thus, in clinical research, questionnaires that include multiple items addressing one or more underlying constructs have proved the most trustworthy approach to measuring aspects of health status, including symptoms, functional status and quality of life[2].

With the growing emphasis on patient-centered care, major international health policy and regulatory authorities have recognized the importance of evaluating PROMs in clinical research and, to an increasing extent, in clinical practice[6-9]. Despite the proliferation of PROMs used in research and practice, there remain substantial challenges with interpretation of their results. Users of PROM results – including clinicians, guideline developers and patients – often have no intuitive notion whether an apparent treatment effect is trivial in magnitude, small but important, moderate or large. To address this problem, researchers developed the concept of the minimal important difference (MID): the smallest change – either positive or negative – in an outcome that patients perceive as important[10].

There are two commonly used approaches for determining the MID: anchor-based and distribution-based methods[11]. Distribution-based methods rely solely on the statistical characteristics of the study sample (e.g. standard error of measurement or standard deviation of PROM scores) and do not reflect the patient's perspective, severely limiting their usefulness in interpretating results. Investigators using the anchor-based approach choose an independent interpretable measure as an external criterion or *anchor* such as transition ratings. The MID is calculated by associating the change in the PROM with an improvement or deterioration captured by the anchor.

In the last three decades, the number of published studies providing anchor-based MID estimates for PROMs has grown rapidly (**Figure 6**), and MIDs are increasingly being used in trials, meta-analysis and guidelines to enhance the interpretability of PROMs. In response to this continually expanding field, our research group has developed resources for facilitating the identification and appraisal of anchor-based MIDs. Briefly, we have conducted a systematic survey to develop an inventory of all published studies empirically estimating MIDs for PROMs in the medical literature (Submitted Dec 2018 to the BMJ), and created (Submitted Dec 2018 to the BMJ) and applied a novel instrument to evaluate the credibility of all 3389 MID estimates for 358 PROMs from 338 studies in our inventory. When summarizing MIDs and evaluating their credibility, we encountered challenges and identified opportunities to improve current methodological standards.
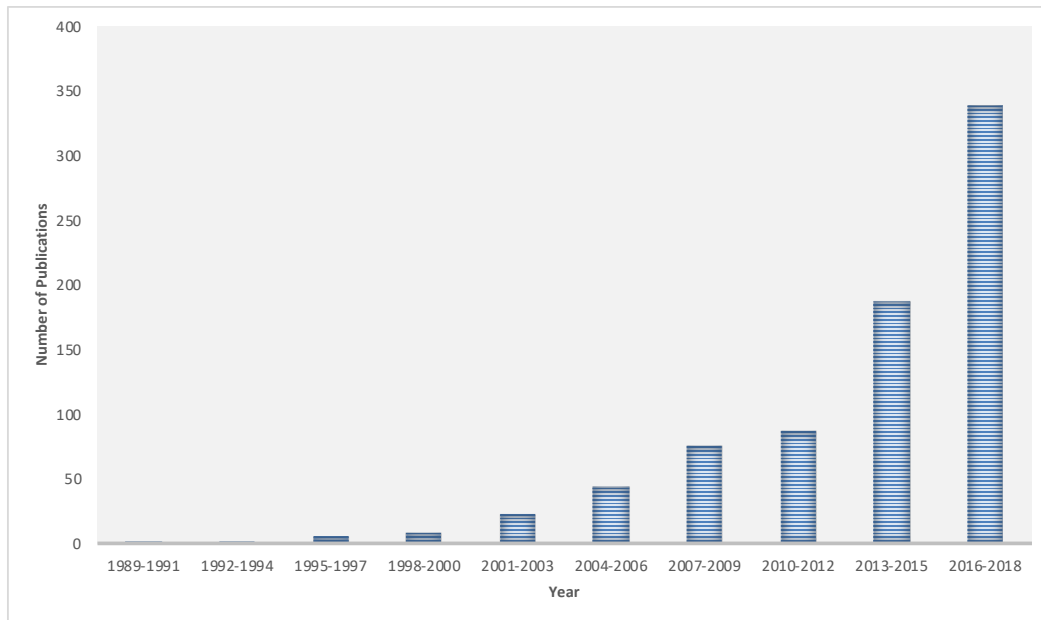
Figure 6. Number of published anchor-based MID estimation studies from 1989 up to October 2018 during each 3-year stratum

According to the Web of Science, the original paper describing the MID[10] has been cited over 2,200 times; in 2017, 28 years after its publication, it was cited 175 times. Given the widespread recognition of the usefulness of the MID, elucidating the most critical issues of which MID users should be aware may be of considerable value. Key stakeholders who may find these insights useful include committees that develop clinical practice guidelines and formularies, set market access and reimbursement policies, and make regulatory decisions; as well as trialists, systematic reviewers, clinicians and patients. Improved understanding of MID concepts and awareness of common pitfalls in methodology and reporting will better inform the application of MIDs in clinical research, clinical practice and regulatory policy.

1. Lack of a consistent nomenclature

The MID concept was first introduced in the medical literature in 1989[10], labelling the concept the minimal *clinical* important difference. Because this terminology

focused attention on the clinical arena as opposed to patients' experience, the same group of researchers later suggested dropping the word "clinical", and relabelled the concept as the minimal important difference[12,13]. The authors subsequently asserted that the patients who are providing information on aspects of their health status are in the best position to ultimately judge whether a difference in a PROM is or is not important[14]. Many might consider this semantic distinction important, and the universal adoption might add clarity to the discussion.

Indeed, in our systematic survey of 338 anchor-based MID studies, we identified 86 unique terms referring to the MID concept (**Appendix 6**). Most deviations from the original and revised terminology were trivial (e.g. *minimum clinically* important difference), but others were more problematic (e.g. clinically relevant change, minimal patient perceivable deterioration, responder definition improvement).

This profusion of terms is often a semantic matter, but sometimes represents a different concept. For instance, MIDs for improvement may sometimes differ from those of deterioration, and thus being explicit about the direction of the difference (or *change*) may be warranted (e.g. minimal important *improvement*, minimal important *deterioration*)[15,16]. Further, some researchers suggest the usefulness of distinguishing between methods that rely on within-person changes and those that quantify differences between groups – a distinction with which we agree: given the groups can only provide average effects that are likely to differ considerably between individuals, differences between groups is a very poor way to establish an MID[17,18].

Such inconsistencies in terminology add unnecessary complexity to reviewers' task in comprehensively identifying relevant MIDs, requiring meticulous inspection of methodology in individual studies to ensure estimates offered truly reflect the MID.

2. **Multitude of diverse methods for MID estimation – Determining whether the MID actually reflects a small but important difference**

Generally speaking, the methodology underpinning an anchor-based MID relies on two key components: 1) the anchor and 2) the analytical (or estimation) approach. The appropriate use of an anchor for MID estimation requires knowledge regarding the magnitude of difference on the anchor that is small but important to patients. Two-thirds of the MIDs in our inventory were estimated using a transition rating – perhaps not surprising given that transition ratings are easily understandable for both clinicians and patients, and can easily be framed to relate closely to the construct that the PROM is measuring (Submitted Dec 2018 to the BMJ). Investigators must, however, choose the right response option to correspond to the MID (e.g. "a little better" would be a choice much superior to "much better"). Quantifying a change that is small but important to patients on other anchor types such as hemoglobin levels, incontinence episodes, or Crohn's disease activity index are likely to be much more challenging.

Although anchors with a very limited relation to patient function or experience (such as hemoglobin) are very likely poor choices of anchors, there is no consensus on the type of anchor that is best suited for ascertaining MIDs. Moreover, for the same anchor, the threshold defining the MID often differs across studies. Applying our credibility instrument to the studies in our inventory revealed that, in our judgement, anchors reflected a small but important difference in only 43% of MIDs.

Further complicating matters, even after the threshold is set, there are a multitude of statistical approaches to compute the MID, each with its own merits and limitations[11] (Table 7). The different analytical methods will yield different estimates[16,19], and whether one or more represent better choices remains unresolved.

Table 7. Analytical methods reported in the anchor-based MID estimation studies included in the MID inventory

1. Changes within patients over time

The MID is the mean change in PROM scores over time within the subgroup of participants who reported a small but important change (i.e. improvement and/or worsening).

The MID is the median change in PROM scores over time within the subgroup of participants who reported a small but important change.

The MID is defined as the 75th percentile of the distribution of change in PROM scores within the subgroup of participants who reported a small but important change.

The MID is the lower or upper limit of the 95% CI of the mean change in PROM scores over time within the subgroup of participants who reported a small but important change.

The MID is estimated using a regression model (either logistic or linear), in which the dependent variable is the change in PROM score and the independent variable is the value, rating or category on the anchor (e.g., ratings on a global rating of change or score on the anchor instrument) that reflects a small but important change. Alternatively, PROM score at follow-up may be the dependent variable, while the independent variables are the value, rating or category on the anchor, and the baseline PROM score.

The MID is estimated using an ANOVA model, in which the dependent variable is the change in PROM score and the independent variable is the value, rating or category on the anchor (e.g., ratings on a global rating of change or score on the anchor instrument).

The MID is estimated using discriminant function analysis.

The MID is estimated using linkage (or scale-alignment) to estimate the MID.

2. Differences between groups capturing changes within patients over time

The MID is the mean change in PROM scores over time in the participants with a small but important change minus the mean change in PROM scores over time in the participants with no change. This method attempts to correct for the change in rating in the no change group.

The MID is the change in PROM scores over time in the participants in one group minus the mean change in PROM scores over time in the participants in another group. The participants in these groups have a different status on the same condition or disease-related outcome.

3. Differences between patients' PROM scores at one time-point

The MID is the difference in PROM scores between participants who rated themselves compared to another participant, as a little bit better (or a little bit worse) versus participants who rated themselves, compared to another participant, as about the same.

The MID is the difference in PROM scores between participants in groups with a different status on the same condition or disease-related outcome.

The MID is estimated using a regression model (either logistic or linear), where the dependent variable is the PROM score and the independent variable is the value, rating or category on the anchor (e.g., score on the anchor instrument).

4.	**Receiver operating characteristic curve analysis**

Methods selecting optimal cut-point based on lowest overall misclassifications and giving equal weight to sensitivity and specificity

The point of the ROC curve that maximizes the distance to the identity (or chance) line is selected as the optimal MID

The closest-to-(0,1) criterion. The point on the ROC curve closest to (0,1) (upper left corner of the graph)

45-degree tangent line. A -45° tangent line is drawn from (0,1) to (1,0) intersecting the ROC curve, i.e., from the top left corner to the bottom right corner of the graph. The cut-point is identified as the point on the ROC curve that is closest to the -45° tangent line. In other words, this is the cut-point with equal (or almost equal) sensitivity and specificity.

Others

80 % specificity rule. The MID is the best sensitivity for response while still achieving at least 80% specificity.

The MID is the cut-point associated with the optimal likelihood ratio. For example, a likelihood ratio of 10 may be selected.

GROC, global rating of change; MID, minimal important difference; PROM, patient reported outcome measure; ROC, receiver operating characteristic; CI, confidence interval

The recent CONSORT PRO Extension addresses the need for enhanced interpretation of PRO results, and encourages authors to include discussion of an MID or a responder definition in clinical trial reports[20]. This demand for increased MID reporting in trials will require clinical trialists and users of trial data to better understand MID methodology and distinguish between more and less trustworthy MID estimates. Failure to use credible MIDs to provide interpretable estimates of treatment effects measured by PROMs may lead to serious misinterpretations of findings from otherwise well-designed clinical trials and meta-analyses.

3. Inadequate reporting and the need for the development of a reporting standard for better transparency

In addition to appropriate design, conduct and analysis, investigators must also report clear, transparent and trustworthy research findings. In the development of our inventory of MIDs, we found major deficiencies in reporting. The usefulness of the anchor-based approach is critically dependent on the extent to which the PROM and anchor measure the same, similar or related constructs. Thus, perhaps arguably, the single most important aspect of credibility of the MID is the correlation between the PROM and anchor. In our inventory we found that, for 71% of MIDs, authors did not report an associated correlation coefficient.

Further, our confidence in an MID estimate will be lower if the confidence interval around the point estimate is insufficiently narrow; yet, for 56% of MIDs, authors did not report a measure of imprecision. Even simpler issues, such as the number of patients informing the MID, proved unclear for 22% of MIDs; a similar number failed to report the range of the patient reported measurement scale for which the MID was determined. This is problematic because different scales often exist for a single PROM. For example, the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) pain instrument may be rated on a 5-point likert-type scale with items summing to give a possible range of 0 to 20 or as a 0 to 100

visual analogue scale. Moreover, in some studies investigators will transform scores to a different scale (e.g. 0 to 100) than the authors who developed the instrument have reported[21].

Judging the credibility and applicability of MIDs requires complete and accurate reporting: inadequate reporting in MID determination studies will limit the value of these data in clinical research and practice. The use of reporting guidelines for other types of research, such as the CONSORT statement for randomized trials, has resulted in superior reporting[22,23]. In similar fashion, the development of a reporting guideline for anchor-based MID estimation studies will likely improve the completeness and transparency of MID reports and promote higher methodologic standards for robust MID estimation.

Although the MID represents a powerful tool for enhancing the interpretability of PROMs, realizing its full value will require improved understanding and reporting of its measurement fundamentals. Some of the issues we have labeled – in particular, terminology and completeness of reporting, are easily addressed. Others, such as choice of optimal anchors and response options representing a small but important threshold difference, and optimal statistical approaches, are likely to prove more challenging. Empirical investigations in the exploration of factors explaining variability in MIDs may aid in informing the desperately needed harmonization of methods.

Summary Points
- Minimal important differences (MIDs) are increasingly being used in trials, meta-analysis and practice guidelines to enhance the interpretability of patient-reported outcome measures (PROMs)
- The profusion of varying terms representing the MID concept adds unnecessary and problematic complexity to users' task in comprehensively

identifying relevant MIDs, requiring meticulous inspection of methodology to ensure estimates offered truly reflect the MID

- There are a multitude of diverse methods for MID estimation that will yield different estimates and whether one or more represent better choices remains unresolved

- The development of a reporting guideline for anchor-based MID estimation studies is likely to address, to an appreciable extent, many of the problems in current presentation of methods and results in MID studies.

Contributors statement

TD wrote the first draft of the manuscript; TD, ACL, GG interpreted the data analysis and critically revised the manuscript. TD is the guarantor.

Competing interests' statement

All authors have completed the ICMJE uniform disclosure form and declare no support from any organization for the submitted work. There are no other relationships or activities that could appear to have influenced the submitted work.

Ethical approval statement: Not required.

Data sharing statement: No additional data available.

Transparency statement: TD affirms that the manuscript is an honest, accurate, and transparent account of the recommendation being reported; that no important

aspects of the recommendation have been omitted; and that any discrepancies from the recommendation as planned (and, if relevant, registered) have been explained.

Appendix 6. Terms referring to the minimal important difference concept across 338 included studies in our MID inventory

Terms
Clinically important difference; minimum deterioration
Clinically important difference; minimum improvement
Clinically important difference
Clinically important difference for deterioration
Clinically important difference for improvement
Clinically meaningful change
Clinically meaningful improvement
Clinically relevant change
Clinically significant improvement
Greater fatigue minimal clinically important difference
Less fatigue minimal clinically important difference
Minimal clinical important difference
Minimal clinical important difference for negative change
Minimal clinical important difference for positive change
Minimal clinically important change
Minimal clinically important change for improvement
Minimal clinically important change for worsening
Minimal clinically important difference
Minimal clinically important difference for deterioration
Minimal clinically important difference for improvement
Minimal clinically important difference for perceived deterioration
Minimal clinically important difference for perceived improvement
Minimal clinically important difference for worsening
Minimal clinically important difference improvement
Minimal clinically important difference of deterioration
Minimal clinically important difference of improvement
Minimal clinically important difference post treatment
Minimal clinically important difference pre treatment
Minimal clinically important improvement
Minimal clinically important score-difference for deterioration
Minimal clinically important score-difference for improvement
Minimal clinically meaningful difference
Minimal clinically significant difference
Minimal clinically significant improvement
Minimal important change
Minimal important change for improvement

Minimal important deterioration
Minimal important difference
Minimal important difference deterioration
Minimal important difference for deterioration
Minimal important difference for improvement
Minimal important difference for patients indicating deterioration
Minimal important difference for patients indicating improvement
Minimal important difference for those who deteriorated
Minimal important difference for those who improved
Minimal important difference for worsening
Minimal important difference improvement
Minimal important improvement
Minimal patient perceivable deterioration
Minimal patient perceivable improvement
Minimally clinical important difference
Minimally clinically identifiable difference
Minimally clinically important change for deterioration
Minimally clinically important difference
Minimally important change
Minimally important change for deterioration
Minimally important change for improvement
Minimally important change for those who improved
Minimally important change improvement
Minimally important difference
Minimally important difference for deterioration
Minimally important difference for improvement
Minimally important difference for those who deteriorated
Minimally important difference for those who improved
Minimally important difference for worsening
Minimally important difference improved
Minimally important difference improvement
Minimally important difference to reflect deterioration
Minimally important difference worsened
Minimally important difference worsening
Minimally important percent change
Minimum clinical important difference
Minimum clinically important change
Minimum clinically important difference
Minimum clinically important improvement
Minimum clinically significant change
Minimum clinically significant difference

Minimum important change
Minimum important difference
Minimum patient perceivable deterioration
Negative minimal clinically important difference
Negative minimal important difference
Positive minimal clinically important difference
Positive minimal important difference
Responder definition deterioration
Responder definition improvement

References

1. Kamper SJ, Maher CG, Mackay G. Global rating of change scales: a review of strengths and weaknesses and considerations for design. The Journal of manual & manipulative therapy. 2009;17(3):163-170.
2. Bowling A. Just one question: If one question works, why ask several? In: BMJ Publishing Group Ltd; 2005.
3. Manning Jr WG, Newhouse JP, Ware Jr JE. The status of health in demand estimation; or, beyond excellent, good, fair, poor. Economic aspects of health: University of Chicago Press 1982:141-84.
4. McHorney CA, Ware Jr JE, Rogers W, et al. The validity and relative precision of MOS short-and long-form health status scales and Dartmouth COOP charts: results from the Medical Outcomes Study. 1992:MS253-MS65.
5. Ware J, Dewey JJIEJHE. Health status and outcomes assessment tools. 2000;3(3):138-48.
6. US Food and Drug Administration (FDA). Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. Health and Quality of Life Outcomes. 2006;4:1-20.
7. European Medicines Agency Committee for Medicinal Products for Human Use. Appendix 2 to the Guideline on the Evaluation of Anticancer Medicinal Products in Man: The Use of Patient-Reported Outcome (PRO) Measures in Oncology Studies EMA/CHMP/292464/2014. London, England: European Medicines Agency. 2016.
8. Doward LC, Gnanasakthy A, Baker MG. Patient reported outcomes: looking beyond the label claim. Health Qual Life Outcomes. 2010;8:89.
9. Valderas J, Kotzeva A, Espallargues M, et al. The impact of measuring patient-reported outcomes in clinical practice: a systematic review of the literature. 2008;17(2):179-193.
10. Jaeschke R, Singer J, Guyatt GH. Measurement of health status: ascertaining the minimal clinically important difference. Controlled clinical trials. 1989;10(4):407-415.
11. King M. The interpretation of scores from the EORTC quality of life questionnaire QLQ-C30. Quality of life research. 1996;5(6):555-567.
12. Juniper EF, Guyatt GH, Willan A, Griffith LE. Determining a minimal important change in a disease-specific quality of life questionnaire. Journal of clinical epidemiology. 1994;47(1):81-87.
13. Schunemann HJ, Puhan M, Goldstein R, Jaeschke R, Guyatt GH. Measurement properties and interpretability of the Chronic respiratory disease questionnaire (CRQ). Copd. 2005;2(1):81-89.
14. Guyatt G, Montori V, Devereaux PJ, Schunemann H, Bhandari M. Patients at the center: in our practice, and in our use of language. ACP journal club. 2004;140(1):A11-12.

15. King MT. A point of minimal important difference (MID): a critique of terminology and methods. *Expert review of pharmacoeconomics & outcomes research* 2011;11(2):171-84. doi: 10.1586/erp.11.9 [published Online First: 2011/04/12]

16. Mills KA, Naylor JM, Eyles JP, Roos EM, Hunter DJ. Examining the Minimal Important Difference of Patient-reported Outcome Measures for Individuals with Knee Osteoarthritis: A Model Using the Knee Injury and Osteoarthritis Outcome Score. J Rheumatol. 2016;43(2):395-404.

17. De Vet HC, Terwee CB, Mokkink LB, et al. Measurement in medicine: a practical guide: Cambridge University Press 2011.

18. Terwee CB, Roorda LD, Knol DL, et al. Linking measurement error to minimal important change of patient-reported outcomes. 2009;62(10):1062-67.

19. Terwee CB, Roorda LD, Dekker J, et al. Mind the MIC: large variation among populations and methods. J Clin Epidemiol. 2010;63(5):524-534.

20. Calvert M, Blazeby J, Altman DG, et al. Reporting of patient-reported outcomes in randomized trials: The consort pro extension. JAMA. 2013;309(8):814-822.

21. American College of Rheumatology. Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC). 2018; https://www.rheumatology.org/I-Am-A/Rheumatologist/Research/Clinician-Researchers/Western-Ontario-McMaster-Universities-Osteoarthritis-Index-WOMAC. Accessed October 31, 2018.

22. Cobo E, Cortes J, Ribera JM, et al. Effect of using reporting guidelines during peer review on quality of final manuscripts submitted to a biomedical journal: masked randomised trial. Bmj. 2011;343:d6783.

23. Turner L, Shamseer L, Altman DG, et al. Consolidated standards of reporting trials (CONSORT) and the completeness of reporting of randomised controlled trials (RCTs) published in medical journals. The Cochrane database of systematic reviews. 2012;11:Mr000030.

Chapter 5: Application of MIDs in degenerative knee disease outcomes: a systematic review and case study to inform BMJ Rapid Recommendations

Tahira Devji, Gordon H. Guyatt, Lyubov Lytvyn, Romina Brignardello-Petersen, Farid Foroutan, Behnam Sadeghirad, Reed A.C. Siemieniuk, Alonso Carrasco-Labra, Per Olav Vandvik

ABSTRACT

Objectives: To identify the most credible anchor-based minimal important differences (MIDs) for patient important outcomes in patients with degenerative knee disease, and to inform BMJ Rapid Recommendations for arthroscopic surgery versus conservative management

Design: Systematic review.

Outcome measures: Estimates of anchor-based MIDs, and their credibility, for knee symptoms and health-related quality of life (HRQoL).

Data sources: MEDLINE, EMBASE and PsycINFO.

Eligibility criteria: We included original studies documenting the development of anchor-based MIDs for patient-reported outcomes (PROs) reported in randomised controlled trials included in the linked systematic review and meta-analysis and judged by the parallel BMJ Rapid Recommendations panel as critically important for informing their recommendation: measures of pain, function and HRQoL.

Results: 13 studies reported 95 empirically estimated anchor-based MIDs for eight PRO instruments and/or their subdomains that measure knee pain, function or HRQoL. All studies used a transition rating (global rating of change) as the anchor to ascertain the MID. Among PROs with more than one estimated MID, we found wide variation in MID values. Many studies suffered from serious methodological limitations. We identified the following most credible MIDs: Western Ontario and McMaster University Osteoarthritis Index (WOMAC; pain: 12, function: 13), Knee injury and Osteoarthritis Outcome Score (KOOS; pain: 12, activities of daily living: 8) and EuroQol five dimensions Questionnaire (EQ-5D; 0.15).

Conclusions: We were able to distinguish between more and less credible MID estimates and provide best estimates for key instruments that informed evidence presentation in the associated systematic review and judgements made by the Rapid Recommendation panel.

Strengths and limitations of this study

- This is the first systematic review of minimal important differences (MIDs) for patient-reported outcomes measuring pain, function and health-related quality of life in patients with degenerative knee disease.

- We demonstrate how MIDs can inform presentation of findings in systematic reviews, and judgements in guideline development.

- There are no established credibility criteria for MIDs with measurement properties, particularly reliability, that have been formally tested.

- Even applying our credibility criterion of a sufficiently high correlation, the range of MIDs reported was very wide; credibility of the estimates may still be limited.

INTRODUCTION

Degenerative knee disease (osteoarthritis in the knee, which can involve the joint lining and/or menisci) is a chronic, progressively debilitating condition, affecting more than nine million people in the USA.[1] A number of randomised controlled trials (RCTs) have assessed the impact of arthroscopic surgery involving partial meniscectomy, debridement or both in patients with degenerative knee disease. These RCTs have reported effects of arthroscopy on patient-reported outcomes (PROs) of knee pain, function and health-related quality of life (HRQoL), which are critical outcomes in degenerative knee disease trials.[2,3] The RCTs have demonstrated that arthroscopic surgery results in a small improvement in pain and function over the short term, but guidance for clinicians and patients requires determining the importance of these benefits.[4]

Investigators are increasingly relying on PROs as key end points in clinical trials. Although PROs provide patients' experience of the impact of disease and treatment on their health status, challenges in interpreting changes in PRO scores can limit their usefulness in informing patient-centred care.[5] For instance, does a 10 mm reduction in self reported knee pain on a 0–100 mm visual analogue scale (VAS) reflect a trivial difference, a small but important difference, a moderate or even a large effect? The key issue for those making recommendations is how patients value the outcomes: in this case, where in the continuum between trivial and very important will patients place observed improvements in pain and function? The smallest change that patients perceive as important, either beneficial or harmful – the minimal important difference (MID)[6,7] – reflects patients' values and preferences, and can therefore enhance the interpretation of PROs, facilitating understanding of the importance of intervention effects in RCTs.

Establishing the MID for an index instrument requires comparison of instrument scores with another instrument (typically referred to as an anchor) that is itself

interpretable. The most popular approach uses a transition instrument (asking patients whether they have improved or deteriorated, and the magnitude of that improvement or deterioration) as the anchor, and relating change in instrument score to the patients' rating of change in status over time.[8] In this method, patients complete the index instrument on two occasions. On the second occasion, they rate the extent to which they have improved or deteriorated; it is this transition rating that provides the anchor. Typically, patients who have experienced a small but important improvement or deterioration inform the MID estimate.

BMJ Rapid Recommendations[9] is a new series of trustworthy recommendations published in response to potentially practice changing evidence. BMJ Rapid Recommendations panels, as in any guideline, require best current evidence to inform their recommendations, covered by one or more linked systematic reviews.[9-11] Another requirement is appropriate interpretation of the importance of effects when moving from evidence to recommendations – judgements that should reflect patients' values and preferences.[10,11] The panel responsible for creating the second BMJ Rapid Recommendations, addressing the impact of arthroscopic surgery versus conservative management in patients with degenerative knee disease, faced challenges in interpreting the significance of apparent treatment effects on the critical outcomes of interest: pain, function and HRQoL from the linked systematic review.[4,12] To help address this challenge, we conducted an additional linked systematic review to identify the most credible anchor-based MID estimates for the PROs used in trials comparing arthroscopic surgery to conservative management. In this paper, we describe our approach to gathering and interpreting the credibility of MID estimates, and note how our results informed the linked systematic review of treatment effectiveness[4] and the subsequent development of the BMJ Rapid Recommendations.[12]

METHODS

Guideline panel and patient involvement

According to the BMJ Rapid Recommendations process,[9] a guideline panel provided critical oversight to our systematic review addressing MID estimates as well as the linked systematic review of effectiveness. The panel, which included eight content experts and front-line clinicians (three orthopaedic surgeons, one rheumatologist, one epidemiologist, one general practitioner and two physiotherapists), four methodologists (three of whom are also front-line clinicians and general internists) and three patients with lived experience of degenerative knee disease, identified populations, subgroups and outcomes of interest.[4,9] Patients received personal training and support throughout the guideline development process.

Patient values and preferences were incorporated in the guideline process through application of the MIDs from our systematic review of studies in which patients provided ratings of the magnitude of change they had experienced, and whether that change was trivial, small but important, or larger. Patients also led the interpretation of the results in the guideline panel based on their assessment of typical patient values and preferences, as well as the variation in values between patients.

Literature search and study identification

We updated our search from a systematic review of anchor-based MIDs[13] that identified articles from 1989 up to 13 April 2015 (the MID concept was first introduced into the medical literature in 1989[8]) using MEDLINE, EMBASE and PsycINFO from 2 February 2015 to 15 September 2016.[8] For the update of our initial search, we added filters for the specific PROs assessed in RCTs included in the linked systematic review and meta-analysis addressing benefits and harms of arthroscopy that informed the guideline panel in making their recommendation.[4] There were no restrictions on language. **Appendix 7** presents the search strategy for MEDLINE, which we adapted for each of the selected databases.

Study selection

The parallel BMJ Rapid Recommendations panel identified pain, function and HRQoL as key patient-important outcomes in the management of degenerative knee disease.[4] We included original reports of studies that empirically estimated an anchor-based MID in patients with degenerative arthritis of the knee for PRO measures that informed the systematic review and meta-analysis of treatment effects for the Rapid Recommendation, that is, outcomes included in the eligible randomised trials.[4] Studies comparing the results of the PRO instrument to an independent standard (the anchor), irrespective of the interpretability or the quality of the anchor, were eligible. Two pairs of reviewers performed title and abstract and full-text screening independently and in duplicate. All studies included by either reviewer in the title and abstract stage were screened in full text. Reviewers resolved disagreements at the full text screening stage through discussion.

Data abstraction

Two pairs of reviewers independently extracted data from eligible studies in duplicate using a standardized pilot tested spreadsheet including the following: first author; publication year; country; participant demographics, including age, sex, condition under investigation; characteristics of the PRO, such as type (generic vs specific), domain(s) and construct(s) captured by the instrument; details pertaining to the method(s) of MID estimation, including number of participants used to estimate the MID, duration of follow-up from baseline, characteristics of the anchor, analysis method (mean change vs receiver operating characteristic (ROC) curves), and correlation between the anchor and PRO scores). We abstracted and report only MIDs for improvement, expressed as absolute estimates, along with the associated 95% CI. We did not include estimates in which the estimated MID was reported as a deterioration.

Credibility assessment

We defined credibility as the extent to which the design and conduct of studies measuring MIDs are likely to have protected against misleading estimates.[13] Although there are numerous established risk of bias and quality grading instruments for use in systematic reviews, none are suited to assess the credibility of studies estimating an MID. We dealt with credibility by focusing on a single criterion that is clearly related to credibility and can be ascertained without judgement: the correlation between change in the index PRO under consideration and the global rating of change that constitutes the anchor. Our threshold for an acceptable correlation was 0.4 or greater.[14–16]

Synthesis of results

We summarised the MID estimates, along with intervention, population characteristics and characteristics of the anchor. We provided the systematic review team with the median, minimum and maximum values across the range of plausible trustworthy MID estimates generated from the eligible studies for the PROs of interest. We pooled the estimates using inverse variance weights and a random-effects model.

To explore potential heterogeneity in MID estimates across studies, we conducted subgroup analyses for possible effect modifiers when we identified at least two studies or two cohorts within studies for each subgroup class (for instance, for nature of intervention, we required at least two surgical cohorts and two nonsurgical cohorts). We considered a number of factors plausibly associated with credibility of estimates including the anchor estimate coming from the patients and interpretable to the patient and clinician, precision around the estimate, whether the anchor represents a minimal change, and the length of time between the initial visit and follow-up. The required number of cohorts in each subgroup class was available for only the last of these. We also performed subgroup analyses comparing MIDs estimated in patients undergoing surgical intervention versus

those receiving conservative management, and in those using ROC curve analysis versus mean change methods. When more than one MID derived from a single study or cohort was provided, we took the median of the estimates. For instance, in our subgroup analysis exploring the effect of intervention type (surgical or nonsurgical) on MID, when authors provided data for more than one time point, we used the median of the available data. To determine if there was a subgroup effect, we considered a test for interaction p value of <0.05 between the proposed variables and the MID to be significant.

STATA software V.12.0 provided software for all analyses.

**Practical application of MID estimates in BMJ Rapid Recommendations development**

Three content experts from the guideline panel with clinical experience with the measures participated in our systematic review of MID estimates, ensuring applicability to the process of developing the recommendations. We applied the MID estimates identified as credible from our review in the evidence summary presented in the linked systematic review addressing treatment effectiveness that informed the BMJ Rapid Recommendations panel in their development of recommendations.[4,12]

The panel used the MID estimates in two ways. One was to intuitively relate the MID estimates to the magnitude of the effect (the smaller the effect in relation to the MID, the less important the effect). The second was to inform statistical techniques to estimate the proportion of patients in intervention and control groups that improved more than the MID, calculate a risk difference on the basis of these results and pool risk differences across studies.[17] We performed sensitivity analyses using the minimum and maximum MIDs across the range of credible estimates for each PRO to test the robustness of our findings.

RESULTS

We screened 4730 unique citations, of which 1716 were judged potentially eligible on review of titles and abstracts, and 15 deemed eligible on full-text review (**Figure 7**). Two[18,19] of the 15 eligible publications provided secondary reports of the same patients included in earlier empirical studies[20,21] estimating MIDs for the same PRO measures. We used both sets of reports to obtain all relevant data for our review.
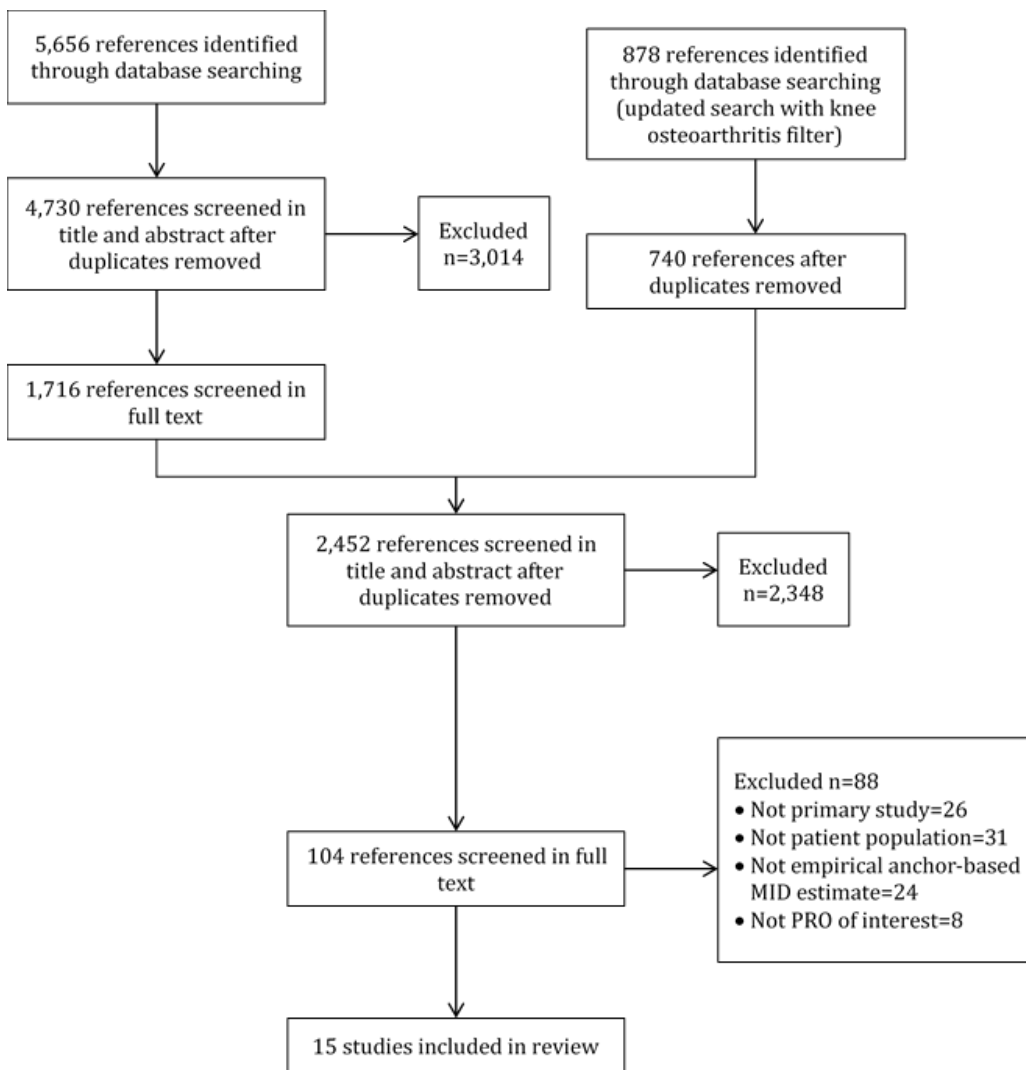
MID, minimally important difference; PRO, patient-reported outcome

Figure 7. Study flow diagram

Table 7 presents the study characteristics. Thirteen studies reported anchor-based MIDs for eight candidate PROs and/or their domains assessing knee pain, function or HRQoL. All studies used a transition rating (a global rating of change) as the anchor to ascertain the MID. The number of patients informing the estimation of the MID ranged from 31 to 497. **Table 7** highlights the studies from which the credible MIDs (those with a correlation of 0.4 or greater between change in the index instrument and the global rating of change) were drawn. Content experts confirmed that the range of patients and treatments included in the final selection of MIDs was satisfactory to inform MIDs for the population, intervention and comparator addressed by the recommendation. Characteristics of included studies reporting MIDs for knee symptoms and health-related quality of life.

More than one study provided estimates for six of the PROs, and all studies derived MIDs for more than one PRO or PRO domain. Two studies[14,26] used more than one anchor to estimate MIDs for the same PRO. Two studies[20,29] estimated MIDs for multiple cohorts of patients and reported the estimates separately. Follow-up duration ranged from 20 days to 24 months. Three studies[14,25,29] estimated MIDs for more than one length of follow-up. Investigators used ROC curves to calculate the MID in one study,[27] mean change methods in nine studies[21–26,28,30,31] and both approaches in three studies.[14,20,29] Altogether, 13 unique studies included in our review reported a total of 95 empirically estimated anchor-based MIDs.

In 20 instances, the correlation between the anchor and the PRO for which the MID was estimated was <0.4. Nine studies[21–23,25–28,30,31] providing 21 MIDs did not provide correlation coefficients. We deemed these 41 estimates not trustworthy and thus did not include them in the plausible range of MIDs. For these reasons, we were unable to present credible MIDs for the VAS pain and 36 item Short Form Survey (SF-36) bodily pain and physical function domains.

Table 8. Characteristics of included studies reporting MIDs for knee symptoms and health-related quality of life

Study	Disease/condition	Intervention	Instrument/scale (abbreviated name)	Score range	Construct(s) measured	Anchor
Angst et al. 2002[22]	Knee/Hip OA	Comprehensive rehabilitation intervention	WOMAC	0 to 100*	Pain, function, stiffness	GROC (5-point)
Bellamy et al. 2015[23]	Knee/Hip OA	NSAIDs	WOMAC	0 to 100†	Pain, function, stiffness	GROC 2-step approach (4-point)
Browne et al. 2010[24]	NR	TKA	EQ-5D	-0.4 to 1	HRQoL	GROC (5-point)
Escobar et al. 2007[25]	Knee OA	TKA	WOMAC	0 to 100*	Pain, function, stiffness	GROC (5-point)
			SF-36	0 to 100	HRQoL	
Escobar et al. 2013/2014[18,20]	Knee OA	TKA	WOMAC†	0 to 100	Pain, Function	Transition rating for pain and function (5-point)
Ornetti et al. 2011[26]	Knee OA	NSAIDs	WOMAC	0 to 100*	Function	GROC (4-point)
Mills et al. 2016[14]	Knee OA	Osteoarthritis chronic care program‖	KOOS	0 to 100	Pain, ADL, HRQoL	2 anchors: Walking and Knee Health (7-point)
Monticone et al. 2013[27]	Knee OA	Rehabilitation following TKA	KOOS§	0 to 100	Pain, symptoms, ADL, Sport/Rec, HRQoL	GROC (5-point)
Terwee et al. 2009[28]	Knee/Hip OA	TKA	WOMAC¶	0 to 20* 0 to 68*	Pain Function	Transition rating for pain and function (15-point)
Terwee et al. 2010[29]	[a]Patients who reported knee pain in the past 12 months	[a]Usual care	WOMAC	0 to 100	Pain, function	Transition rating for pain (6-point)
	[b]Patients who visited their GP with a new episode of	[b]Usual care	WOMAC	0 to 100	Pain, function	Transition rating for knee

	nontraumatic knee complaints					complaints (6-point)
	[c]Patients who visited their GP with a new episode of knee complaints	[c]Usual care	WOMAC	0 to 100	Pain, function	Transition rating for knee complaints (5-point)
	[d]Knee OA	[d]Behavioural-graded activity or usual care	WOMAC	0 to 100	Pain, function	Transition rating for knee complaints (8-point)
	[e]Knee OA	[e]TKA	WOMAC	0 to 100	Pain, function	Transition rating for pain and function (15-point)
Tubach et al. 2005/2006[19,21]	Knee OA	NSAIDs	VAS pain	0 to 100*	Pain	Response to treatment (5-point)
			WOMAC	0 to 100*	Function	
Tubach et al. 2012[30]	Knee and/or hip OA	NSAIDs	Pain NRS	0 to 10*	Pain	GROC 2-step approach (4-point)
Walters et al. 2005[31]	Knee OA patients recruited from rheumatology clinics and those assessed pre-operatively for TKA	NR	EQ-5D	-0.59 to 1	HRQoL	GROC (5-point)

Bold text were those that provided credible MIDs (those with a correlation of 0.4 or greater between change in the index instrument and the global rating of change)

[a-e]Terwee et al[29] reported on five cohorts of patients from different studies. a=cohort 1; b=cohort 2; c=cohort 3; d=cohort 4; e=cohort 5

*Higher scores on the PRO scale represent a worse outcome

‡Surgical therapy: debridement, shaving, drilling, autologous chondrocyte implantation, abrasion arthroplasty, microfracture, and cell therapy

‖Osteoarthritis chronic care program: multidisciplinary nonsurgical management strategy

¶Dutch

§Italian

†Spanish

ADL, activities of daily living; EQ-5D, EuroQol five dimensions Questionnaire; GP, general practitioner; GROC, global rating of change; HRQoL, health-related quality of life; KOOS, Knee injury and Osteoarthritis Outcome Score; MID, minimal important difference; NR, not reported; NRS, numerical rating scale; NSAID, non-steroidal anti-inflammatory drugs; OA, osteoarthritis; OKS, Oxford Knee Score; PRO, patient-reported outcome; SF-36, 36-Item Short Form Survey; TKA, total knee arthroplasty; VAS, visual analogue scale; WOMAC, Western Ontario and McMaster University Osteoarthritis Indexs

Table 8 presents the median absolute MID estimate for the Western Ontario and McMaster University Osteoarthritis Index (WOMAC) and Knee injury and Osteoarthritis Outcome Score (KOOS) pain and function domains, and EuroQol five dimensions Questionnaire (EQ-5D), along with the minimum and maximum values across the range of plausible trustworthy estimates (ie, those in which correlations were 0.4 or greater). Among PROs with more than one estimated MID, even among those with correlations of 0.4 or greater, we found wide variation in MID values. **Appendix 8** presents the MID estimates, as well as details regarding MID estimation for each PRO measure. The content experts confirmed that the MID thresholds generated were consistent with their impressions from use of the instruments in clinical practice.

Table 9. Summary of the range of plausible credible MIDs for improvement for PRO measures used to inform the SR of treatment effects

| PRO Instrument/domain (score range)	Absolute MID*	Minimum	Maximum
WOMAC			
Pain (0-100)	12	2	30
Function (0-100)	13	3	34
KOOS			
Pain (0-100)	12	4	20
ADL (0-100)	8	3	9
EQ-5D (-0.59-1)	0.15	NR	NR

MID estimates are presented as positive values, regardless of the direction of change
*Median MID estimate
ADL, activities of daily living; EQ-5D, EuroQol five dimensions Questionnaire; KOOS, Knee injury and Osteoarthritis Outcome Score; MID, minimal important difference; NR, not reported; PRO, patient-reported outcome; WOMAC, Western Ontario and McMaster University Osteoarthritis Index
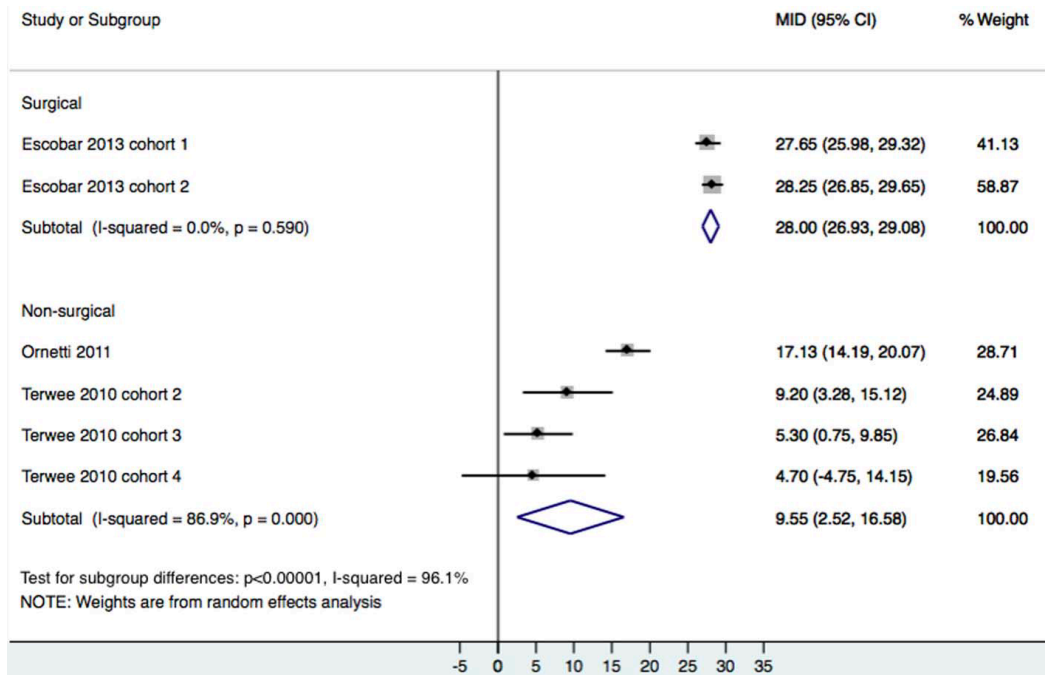
We only performed subgroup analyses exploring potential sources of heterogeneity for the WOMAC pain and function domains, as estimates for the KOOS pain and activities of daily living, and EQ-5D came from a single study. Type of intervention (ie, total knee arthroplasty (TKA) vs conservative management) was significantly

associated with magnitude of the MID for WOMAC pain (p<0.00001; **Figure 8**) and function (p<0.00001; **Figure 9**). For pain, the weighted pooled MID for TKA was 25 (95% CI 24 to 27) in TKA and for conservative management 8 (95% CI 3 to 13). For function, the weighted pooled MID for TKA was 28 (95% CI 27 to 29), and for conservative management 19 (95% CI 3 to 17). We found no association between the hiatus between initial and follow-up visits, nor between the analytic method (ROC or mean change) and the MID.
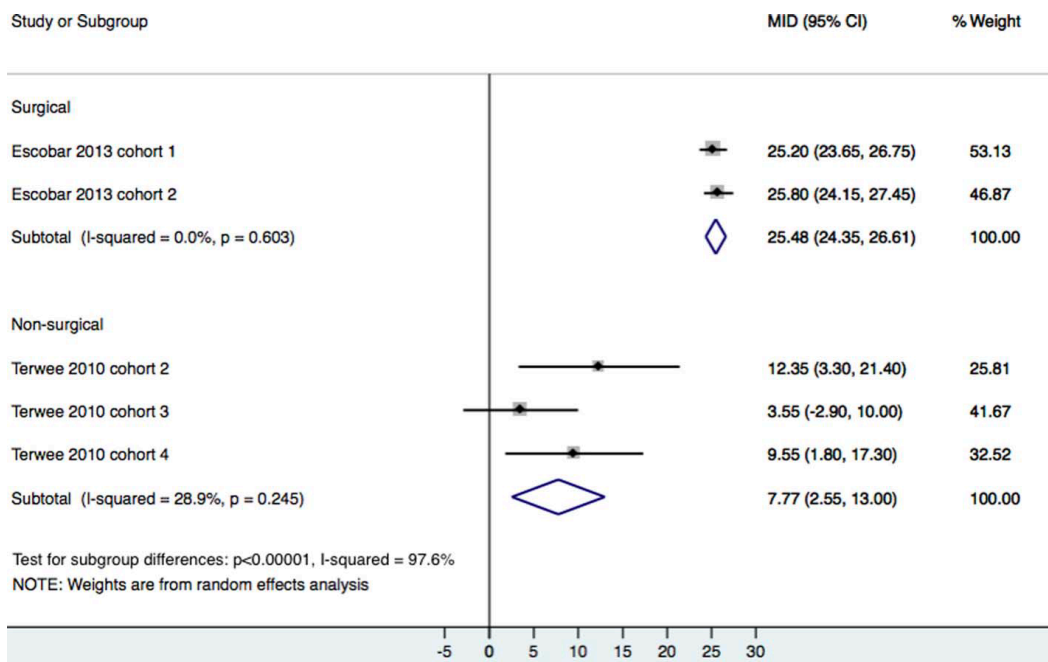
MID, minimally important difference; WOMAC, Western Ontario and McMaster University Osteoarthritis Index

Figure 8. Subgroup analysis for WOMAC pain by intervention type

Study or Subgroup

Study or Subgroup	MID (95% CI)	% Weight
Surgical		
Escobar 2013 cohort 1	27.65 (25.98, 29.32)	41.13
Escobar 2013 cohort 2	28.25 (26.85, 29.65)	58.87
Subtotal (I-squared = 0.0%, p = 0.590)	28.00 (26.93, 29.08)	100.00
Non-surgical		
Ornetti 2011	17.13 (14.19, 20.07)	28.71
Terwee 2010 cohort 2	9.20 (3.28, 15.12)	24.89
Terwee 2010 cohort 3	5.30 (0.75, 9.85)	26.84
Terwee 2010 cohort 4	4.70 (-4.75, 14.15)	19.56
Subtotal (I-squared = 86.9%, p = 0.000)	9.55 (2.52, 16.58)	100.00

Test for subgroup differences: p<0.00001, I-squared = 96.1%
NOTE: Weights are from random effects analysis

MID, minimally important difference; WOMAC, Western Ontario and McMaster University Osteoarthritis Index

Figure 9. Subgroup analysis for WOMAC function by intervention type

Incorporation into the systematic review informing BMJ Rapid Recommendations

The results of this study informed both the systematic review of treatment effects and the Rapid Recommendations panel in their development of recommendations for arthroscopic surgery versus conservative management in patients with degenerative knee disease.[4,12] The panel members reviewed the evidence summary (GRADE Summary of Findings table – **Table 9**) from the systematic review with data addressing pain, function, HRQoL and adverse events; they discussed recommendations through teleconferences.

Table 10. Summary of findings for the outcome of short-term pain presented to *BMJ* Rapid Recommendations Panel

Outcome Timeframe	Study results and measurements	Absolute effect estimates		Certainty in effect estimates (Quality of evidence)	Summary
		Conservative management	Arthroscopy		
Short term					
Pain (difference in change from baseline) 3 months	Measured by: Different instruments converted to scale of index instrument (KOOS pain sub scale) Scale: 0-100 High better, Minimally important difference 12) Data from 1231 patients in 10 studies Follow up 3 months	**15.0 points (Mean)**	**20.0 points (Mean)** **Difference: Mean Difference 5.4 more** (CI 95% 1.9 more - 8.8 more)	**High**	On average, knee arthroscopy results in very small extra reduction in pain scores when compared to control
Pain (difference in patients who achieve a change higher than the MID) 3 months	Data from 1102 patients in 9 studies Follow up 3 months	**669 per 1000**	**793 per 1000** **Difference: 124 more per 1000** (CI 95% 44 more - 204 more)	**High**	Knee arthroscopy increases the number of patients with an important reduction in short-term pain by approximately 12 in 100
Long term					
Pain (difference in change from baseline) 1-2 years	Measured by: Different instruments converted to scale of index instrument (KOOS pain sub scale-Minimally Important Difference 12) Scale: 0-100 High better Based on data from 1097 patients in 8 studies Follow up 2 years	**19.0 points (Mean)**	**22.0 points (Mean)** **Difference: Mean Difference 3.13 more** (CI 95% 0.17 fewer - 6.43 more)	**High**	On average, knee arthroscopy results in no difference, or a very small reduction, in pain

MID, minimally important difference; MD, mean difference

The Summary of Findings for short-term and long term outcome of pain (**Table 9**) exemplifies how the MID for the KOOS pain domain informed this PRO

assessment. Although results from the systematic review favoured arthroscopic surgery in the short term, the estimate of this difference (5.4 points) and its CI (1.9 to 8.8) show magnitudes of effects less than the MID of 12 points established for the index (KOOS) instrument. The systematic review found – by dichotomising outcomes – that 12.4% (95% CI 4.4% to 20.4%) more patients receiving arthroscopy reported a small but important benefit in pain or function at 3 months, which was no longer apparent at 1 year. Sensitivity analyses using the upper and lower estimates across the range of credible MIDs for each instrument, and based on the standardised mean difference (SMD), revealed similar results. The risk difference when using the lowest value of the range was 10.5% (95% CI 4.3% to 16.7%) and when using the highest value of the range it was 11.3% (95% CI 2.9% to 19.7%). The risk difference based on the SMD was 9% (95% CI 1.7% to 15.7%).

The panel was confident in concluding that any benefit from arthroscopic knee surgery is small or very small, and is less important than the burden and transient pain and limitation associated with the arthroscopy procedure itself. The information provided by the MID informed these judgements, which motivated the panel's decision to make a strong recommendation against arthroscopy in patients with degenerative knee disease.

DISCUSSION

Principal findings

In this review, we identified 13 studies reporting MIDs for eight PROs and/or domains measuring knee pain, function and HRQoL in patients with degenerative knee disease, yielding 95 empirically estimated anchor-based MIDs. Investigators used the same anchor-based approach, relying on transition ratings (global ratings of change).[8] For the majority of the PROs, more than one study provided MID estimates, and did so at more than one duration of follow-up, using different

anchors, and using various analytic methods, resulting in multiple estimates for the same PRO.

MID estimates for the same instrument varied widely across all estimates, as well as when restricted to studies meeting our credibility criterion of a correlation of 0.4 or greater between change in the index instrument and the transition rating. Including only MIDs generated from data meeting this criterion, we were able to provide a range of plausible trustworthy estimates for PROs identified as critical outcomes to inform the systematic review of treatment effects and rapid recommendation (**Table 8**). The systematic review team used the most credible MIDs identified in our review to contextualise mean differences and calculate risk differences, and to conduct appropriate sensitivity analyses, expressing the proportion of patients achieving improvements greater than the MID (**Table 9**).

Strengths and limitations

Our study represents the first comprehensive synthesis and evaluation of anchor-based MIDs for self-reported patient-important outcomes commonly assessed in RCTs of degenerative knee disease. We undertook a transparent approach to appraising the credibility of MIDs that allowed identification of the highest credibility MIDs for each instrument. Both the systematic review team and the Rapid Recommendation panel found the MIDs useful in understanding and presentation of the evidence; in particular, the recommendation panel, to a considerable extent, based recommendations on our findings.

This review also has limitations. There are no established credibility criteria for MIDs with measurement properties, particularly reliability, that have been formally tested. We have therefore focused on a single criterion that is indisputably important, and can be ascertained without judgement, and thus without error:

correlations between change in the index instrument and the global rating of change of 0.4 or greater.

Even applying our credibility criterion of a sufficiently high correlation, the range of MIDs reported was very wide (**Table 8**). This raises questions regarding whether the criterion is sufficient – that is, credibility of the estimates may still be limited. The results may, however, reasonably represent a range in which the MID actually lies. We have dealt with this issue by recommending a sensitivity analysis including the full range of plausible MIDs, an approach that the linked systematic review has followed. In future, development and testing of other credibility criteria, and their application in establishing trustworthy MIDs, will strengthen the field.

No MIDs were assessed in the patients of direct interest for the associated systematic review and guideline: patients undergoing knee arthroscopic interventions. Patients included in the eligible studies either underwent major surgery (knee arthroplasty) or non-surgical interventions. Patients did, however, suffer from degenerative knee disease, the condition in which knee arthroscopy is of putative benefit.

For the WOMAC pain and function domains, MIDs estimated in patients undergoing TKA were, on average, appreciably higher than the median MIDs we used as the best estimates. These results suggest that patients undergoing knee arthroplasty, versus those undergoing non-surgical interventions, require a greater degree of change on the PRO measure to consider themselves having an important improvement. In other words, differences in the magnitude of the MID may be related to patient expectations with regard to surgical interventions, as compared with non-surgical or less invasive interventions. The intervention of interest for this Rapid Recommendation – arthroscopy – is, in its invasiveness and immediate consequences, intermediate between non-surgical interventions and total joint

arthroplasty. To the extent that, as a result, our best estimate of the MID underestimates the true MID for arthroscopy, the conclusion in the linked systematic review that the effects of arthroscopy are small or very small is actually strengthened.

One of our PROs, the EQ-5D, had only one MID estimate with a correlation of over 0.4. Moreover, this estimate, 0.15, is inconsistent with evidence from other studies that suggest that 0.15 approximates the entire burden of moderate osteoarthritis, and that the MID for the EQ-5D is appreciably <0.15.[32] For the purpose of the review our work is informing, however, this issue was of minimal concern: the benefits of arthroscopic surgery on quality of life in the short term and long term were not statistically significantly different from patients receiving conservative management, and thus the MID was not needed to further contextualize these results.

Another issue concerns the possible influence of baseline score on the MID.[33] Three of the included studies[18,21,22] in our review reported MIDs for patients stratified according to baseline severity status. Given that MIDs were consistently higher in magnitude with increasing baseline severity, expression of the MID as a relative change may in instances be superior to an absolute difference. A recent report examining the merits of expressing MIDs as relative or absolute estimates in a number of studies suggested, however, that absolute changes generally correlate higher with global change ratings and are simpler to use and interpret.[34]

The following considerations mitigate the concerns regarding the credibility of the MID estimates that guided the panel's recommendation. First, our best estimates of the MID approximate 10% of the instruments' total range, a value that is both intuitive and consistent with MID estimates for other instruments. Second, our best estimates of the MID are consistent with the experience of clinicians who have used

the instruments as part of their clinical practice. Third, estimates for the risk difference in proportion improved with arthroscopy from the sensitivity analyses in the linked systematic review show that using the upper and lower boundaries of the MID that we have suggested, and a value based on the SMD, approximate those using our best estimate of the MID.[4,12]

Implications of the findings for future directions

Our review focused on studies using an anchor-based approach, relying on transition ratings as the anchor, to estimate MIDs; we have highlighted shortcomings in their application. We have focused on a single criterion, correlations of 0.4 and greater, to define credible MIDs. The variability in MIDs generated when this criterion was met suggests residual variation in credibility that warrants further investigation.

Authors have suggested – either explicitly or implicitly, when commenting on strengths and limitations of their studies – criteria for judging credibility of MID estimates emerging from empirical studies. Our group has conducted a systematic survey of such commentaries, and on that basis has developed credibility criteria for studies that define MIDs. Feedback from a wider community will be necessary to establish the robustness, appropriateness and comprehensiveness of these criteria, as well as the empirical studies necessary to establish their reliability.

Given the current uncertainty around MIDs, we recommend that triallists, systematic reviews, guideline panelists and other end users of clinical trial PRO data triangulate their interpretation of these subjective outcomes with additional strategies that complement use of the MID. These include viewing the magnitude of effect in relation to the range of the scale for specific PROs, relying on the experience of clinicians using the instruments in their practice, as well as the use of other summary effect measures (eg, SMD). If interpretations of the results are

consistent across approaches, this will strengthen interpretation of the magnitude of intervention effects.

CONCLUSIONS

The MID has the potential to help interpret the magnitude of treatment effects and thus guide clinical decision making in chronic disease management. This study provides a model for applying the MID concept to aid in the interpretation of evidence, and the formulation of recommendations for clinical practice guidelines.[4,12] Investigators and guideline panelists can use the approaches reported here to make their systematic reviews more informative, and their recommendations more informed, appropriate and useful.

AUTHORS' CONTRIBUTIONS, ACKNOWLEDGMENTS & COMPETING INTERESTS

REFERENCES

1. Lawrence RC, Felson DT, Helmick CG, et al. Estimates of the prevalence of arthritis and other rheumatic conditions in the United States. Part II. Arthritis Rheum 2008;58:26–35.
2. Altman R, Brandt K, Hochberg M, et al. Design and conduct of clinical trials in patients with osteoarthritis: recommendations from a task force of the Osteoarthritis Research Society. Results from a workshop. Osteoarthritis Cartilage 1996;4:217–43.
3. Bellamy N, Kirwan J, Boers M, et al. Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip, and hand osteoarthritis. Consensus development at OMERACT III. J Rheumatol 1997;24:799–802.
4. Brignardello-Petersen R, Guyatt G, Schandelmaier S, et al. Knee arthroscopy versus conservative management in patients with degenerative knee disease: a systematic review. BMJ Open 2017;7: e016114.
5. Boyce MB, Browne JP, Greenhalgh J. The experiences of professionals with using information from patient-reported outcome measures to improve the quality of healthcare: a systematic review of qualitative research. BMJ Qual Saf 2014;23:508–18.
6. Schünemann HJ, Guyatt GH. Commentary—Goodbye M(C)ID! Hello MID, where do you come from? Health Serv Res 2005;40:593–7.
7. Schünemann HJ, Puhan M, Goldstein R, et al. Measurement properties and interpretability of the chronic respiratory disease questionnaire (CRQ). COPD 2005;2:81–9.
8. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Control Clin Trials 1989;10:407–15.
9. Siemieniuk RA, Agoritsas T, Macdonald H, et al. Introduction to BMJ Rapid Recommendations. BMJ 2016;354:i5191.
10. Alonso-Coello P, Oxman AD, Moberg J, et al. GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 2: Clinical practice guidelines. BMJ 2016;353:i2089.
11. Laine C, Taichman DB, Mulrow C. Trustworthy clinical guidelines. Ann Intern Med 2011;154:774–5.
12. Siemieniuk RAC, Harris IA, Agoritsas T, et al. Arthroscopic surgery for degenerative knee arthritis and meniscal tears: a clinical practice guideline. BMJ 2017;257:j1982.
13. Johnston BC, Ebrahim S, Carrasco-Labra A, et al. Minimally important difference estimates and methods: a protocol. BMJ Open 2015;5:e007953.
14. Mills KA, Naylor JM, Eyles JP, et al. Examining the minimal important difference of patient reported outcome measures for individuals with knee

osteoarthritis: a model using the knee injury and osteoarthritis outcome score. J Rheumatol 2016;43:395–404.

15. Revicki D, Hays RD, Cella D, et al. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. J Clin Epidemiol 2008;61:102–9.

16. Turner D, Schünemann HJ, Griffith LE, et al. The minimal detectable change cannot reliably replace the minimal important difference. J Clin Epidemiol 2010;63:28–36.

17. Thorlund K, Walter SD, Johnston BC, et al. Pooling health-related quality of life outcomes in meta-analysis—a tutorial and review of methods for enhancing interpretability. Res Synth Methods 2011;2:188–203.

18. Escobar A, Riddle D. Concordance between important change and acceptable symptom state following knee arthroplasty: the role of baseline scores. Osteoarthritis Cartilage 2014;22: 1107–10.

19. Tubach F, Dougados M, Falissard B, et al. Feeling good rather than feeling better matters more to patients. Arthritis Rheum 2006;55:526–30.

20. Escobar A, Garcia Perez L, Herrera-Espineira C, et al. Total knee replacement; minimal clinically important differences and responders. Osteoarthritis Cartilage 2013;21:2006–12.

21. Tubach F, Ravaud P, Baron G, 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.

22. Angst F, Aeschlimann A, Michel BA, et al. Minimal clinically important rehabilitation effects in patients with osteoarthritis of the lower extremities. J Rheumatol 2002;29:131–8.

23. Bellamy N, Hochberg M, Tubach F, et al. Development of multinational definitions of minimal clinically important improvement and patient acceptable symptomatic state in osteoarthritis. Arthritis Care Res 2015;67:972–80.

24. Browne JP, van der Meulen JH, Lewsey JD, et al. Mathematical coupling may account for the association between baseline severity and minimally important difference values. J Clin Epidemiol 2010;63:865–74.

25. Escobar A, Quintana JM, Bilbao A, et al. Responsiveness and clinically important differences for the WOMAC and SF-36 after total knee replacement. Osteoarthritis Cartilage 2007;15:273–80.

26. Ornetti P, Dougados M, Paternotte S, et al. Validation of a numerical rating scale to assess functional impairment in hip and knee osteoarthritis: comparison with the WOMAC function scale. Ann Rheumatic Dis 2011;70:740–6.

27. Monticone M, Ferrante S, Salvaderi S, et al. Responsiveness and minimal important changes for the knee injury and osteoarthritis outcome score in subjects undergoing rehabilitation after total knee arthroplasty. Am J Phys Med Rehabil 2013;92:864–70.

28. Terwee CB, Roorda LD, Knol DL, et al. Linking measurement error to minimal important change of patient-reported outcomes. J Clin Epidemiol 2009;62:1062–7.

29. Terwee CB, Roorda LD, Dekker J, et al. Mind the MIC: large variation among populations and methods. J Clin Epidemiol 2010;63:524–34.
30. Tubach F, Ravaud P, Martin-Mola E, et al. Minimum clinically important improvement and patient acceptable symptom state in pain and function in rheumatoid arthritis, ankylosing spondylitis, chronic back pain, hand osteoarthritis, and hip and knee osteoarthritis: results from a prospective multinational study. Arthritis Care Res (Hoboken) 2012;64:1699–707.
31. Walters SJ, Brazier JE. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. Qual Life Res 2005;14:1523–32.
32. Sullivan PW, Ghushchyan V. Preference-based EQ-5D index scores for chronic conditions in the United States. Med Decis Making 2006;26:410–20.
33. Rouquette A, Blanchin M, Sébille V, et al. The minimal clinically important difference determined using item response theory models: an attempt to solve the issue of the association with baseline score. J Clin Epidemiol 2014;67:433–40.
34. Zhang Y, Zhang S, Thabane L, et al. Although not consistently superior, the absolute approach to framing the minimally important difference has advantages over the relative approach. J Clin Epidemiol 2015;68:888–94.

Appendix 7. Search Strategies for MEDLINE

Original search (1989 to April 13th, 2015)

Minimally Important Difference Medline Search Strategy

35. (clinical* important difference? Or clinical* meaningful difference? Or clinical* meaningful improvement? Or clinical* relevant mean difference? Or clinical* significant change? Or clinical* significant difference? Or clinical* important improvement? Or clinical* meaningful change? Or mcid or minim* clinical* important or minim* clinical* detectable or minim* clinical* significant or minim* detectable difference? Or minim* important change? Or minim* important difference? Or smallest real difference? Or subjectively significant difference?).tw.

36. "Quality of Life"/

37. "outcome assessment(health care)"/ or treatment outcome/ or treatment failure/

38. exp pain/

39. exp disease attributes/ or exp "signs and symptoms"/

40. or/2-5

41. 1 and 6

42. health status indicators/ or "severity of illness index"/ or sickness impact profile/ or interviews as topic/ or questionnaires/ or self report/

43. Pain Measurement/

44. patient satisfaction/ or patient preference/

45. or/8-10

46. 7 and 11

47. limit 12 to yr="1989 –Current"

48. (quality of life or life 195atisfy??? Or hrqol or hrql).mp.

49. (assessment? Outcome? Or measure? Outcome? Or outcome? Studies or outcome? Study or outcome? Assessment? Or outcome? management or outcome? measure* or outcome? research or patient? outcome? or research outcome? or studies outcome? or study outcome? or therap* outcome? or treatment outcome? or treatment failure?).mp.

50. pain????.mp.

51. ((activity or sever* or course) adj3 (disease or disabilit* or symptom*)).mp.

52. or/14-17

53. 1 and 18

54. (questionnaire? or instrument? or interview? or inventor* or test??? or scale? or subscale? or survey? or index?? or indices or form? or score? or measurement?).mp.

55. (patient? rating? or subject* report? or subject* rating? or self report* or self evaluation? or self appraisal? or self assess* or self rating? or self rated).mp.

56. (patient? report* or patient? observ* or patient? satisf*).mp.

57. anchor base??.mp.
58. or/20-23
59. 19 and 24
60. limit 25 to yr="1989 -Current"
61. 13 or 26

Search Update with Degenerative Knee Disease Filter (February 02 2015 to September 15 2016)
62. Knee.mp.
63. (WOMAC or WOMUOI).mp.
64. KOOS.mp.
65. KOS?ADL.mp.
66. IKDC.mp.
67. Lysholm.mp.
68. OKS.mp.
69. (Activity Rating Scale or ARS).mp.
70. (Tegner Activity Score or TAS).mp.
71. (Visual Analogue Scale or VAS).mp.
72. (Short form or SF?36 or SF?12 or SF?8).mp.
73. (RAND?36 or RAND?12 or RAND?8).mp.
74. (Medical outcomes study or MOS).mp.
75. (EuroQol or EQ?5D).mp.
76. (Health Assessment Questionnaire or HAQ or HAQ?DI).mp.
77. Lequesne.mp.
78. Numerical Rating Scale.mp.
79. Number of painful days.mp.
80. Arthritis.mp.
81. Osteoarthritis.mp.
82. (AIMS or AIMS?2).mp.
83. McGill Pain.mp.
84. ASES.mp.
85. (Schmerzempfindungsskala or SES or Pain Experience Scale).mp.
86. (Pain disability index or PDI).mp.
87. MACTAR.mp.
88. (Brief Pain Inventory or Wisconsin Brief Pain Questionnaire).mp.
89. (Western Ontario Meniscal Evaluation Tool or WOMET).mp.
90. 15?D.mp.
91. (Health Utilities Index or HUI?2 or HUI?3).mp.
92. or/28-57
93. 27 and 58
94. limit 59 to ed=20150201-20160915

Appendix 8. Summary of MID estimates

PRO Instrument	Study	No. of patients	Absolute MID Estimate (SD or 95%CI)	Correlation coefficient between anchor and PRO	Follow up (months)	Method of analysis
WOMAC						
Pain	Terwee 2010[b]	NR	11.8 (7.9 to 15.7)	11.8	12	MC[b]
	Terwee 2010[c]	NR	3.5 (-1.3 to 8.3)	3.5	6	MC[b]
	Terwee 2010[d]	NR	7.9 (1.6 to 14.2)	7.9	9	MC[b]
	Terwee 2010[d]	NR	15.7 (6.5 to 24.9)	15.7	15	MC[b]
	Terwee 2010[b]	NR	12.9 (-1.3 to 27.1)	12.9	12	ROC
	Terwee 2010[c]	NR	3.6 (-4.5 to 11.7)	3.6	6	ROC
	Terwee 2010[d]	NR	2.2 (-11.4 to 15.8)	2.2	9	ROC
	Terwee 2010[d]	NR	11.2 (5 to 17.4)	11.2	15	ROC
	Escobar 2013	209	29.9 (27.1 to 32.6)	29.9	12	MC[b]
	Escobar 2013	207	28.1 (25.1 to 31)	28.1	12	MC[b]
	Escobar 2013	415	20.5 (20.2 to 20.9)	20.5	12	ROC
	Escobar 2013	497	23.5 (23.1 to 23.8)	23.5	12	ROC
Function	Terwee 2010[b]	NR	9.5 (6.5 to 13.4)	9.5	12	MC[b]
	Terwee 2010[c]	NR	5.7 (1.8 to 9.6)	5.7	3	MC[b]
	Terwee 2010[c]	NR	3.3 (0.1 to 6.5)	3.3	6	MC[b]
	Terwee 2010[c]	NR	6.9 (1.8 to 12)	6.9	12	MC[b]
	Terwee 2010[d]	NR	4.7 (0.5 to 8.9)	4.7	9	MC[b]
	Terwee 2010[d]	NR	4.7 (-5.2 to 14.6)	4.7	15	MC[b]
	Terwee 2010[b]	NR	8.9 (0.5 to 17.3)	8.9	12	ROC
	Terwee 2010[c]	NR	6.8 (-1.7 to 15.2)	6.8	3	ROC
	Terwee 2010[c]	NR	3.5 (-0.5 to 7.5)	3.5	6	ROC
	Terwee 2010[c]	NR	4.9 (-10.5 to 20.3)	4.9	12	ROC
	Terwee 2010[d]	NR	3.1 (-9.2 to 15.4)	3.1	9	ROC
	Terwee 2010[d]	NR	5.7 (-3.3 to 14.7)	5.7	15	ROC
	Ornetti 2011	NR	17.1 (14.2 to 20.1)	17.13	1	MC[b]
	Escobar 2013	221	31.1 (28.3 to 33.9)	31.1	12	MC[b]
	Escobar 2013	237	33.5 (30.9 to 36)	33.5	12	MC[b]
	Escobar 2013	415	24.2 (23.6 to 24.7)	24.2	12	ROC
	Escobar 2013	497	23 (22.7 to 23.2)	23	12	ROC

KOOS

Pain	Mills 2016[a]	184	11.5	0.453	6	ROC - Youden
	Mills 2016[a]	184	18	0.453	6	ROC - 80%
	Mills 2016[a]	35	20.0 (8.9 to 31.2)	0.453	6	MC[b]
	Mills 2016[a]	67	15.0 (2.1 to 27.8)	0.453	6	MC[a]
	Mills 2016[b]	184	11.5	0.455	6	ROC - Youden
	Mills 2016[b]	184	17	0.455	6	ROC - 80%
	Mills 2016[b]	34	16.6 (9.9 to 23.2)	0.455	6	MC[b]
	Mills 2016[b]	65	5.6 (-7.7 to 18.9)	0.455	6	MC[a]
	Mills 2016[a]	184	4	0.412	12	ROC - Youden
	Mills 2016[a]	184	13	0.412	12	ROC - 80%
	Mills 2016[a]	35	8.2 (0.7 to 15.9)	0.412	12	MC[b]
	Mills 2016[a]	67	9.3 (-1.9 to 20.4)	0.412	12	MC[a]
	Mills 2016[b]	184	4	0.437	12	ROC - Youden
	Mills 2016[b]	184	12	0.437	12	ROC - 80%
	Mills 2016[b]	34	4.3 (-3.4 to 12.1)	0.437	12	MC[b]
	Mills 2016[b]	65	4.0 (-7.7 to 15.4)	0.437	12	MC[a]
ADL	Mills 2016[a]	184	2.5	0.502	12	ROC - Youden
	Mills 2016[a]	184	-1.5	0.502	12	ROC - 80%
	Mills 2016[a]	35	8.7 (2.9 to 14.2)	0.502	12	MC[b]
	Mills 2016[a]	67	8.9 (0.74 to 17.1)	0.502	12	MC[a]
	Mills 2016[b]	184	2.5	0.471	12	ROC - Youden
	Mills 2016[b]	184	-1.5	0.471	12	ROC - 80%
	Mills 2016[b]	34	8.2 (2.2 to 14.1)	0.471	12	MC[b]
	Mills 2016[b]	65	7.6 (-0.79 to 16.1)	0.471	12	MC[a]

EQ-5D

	Browne 2010	100	0.15	0.42	6	MC[a]

Terwee 2010[a-e]: This study reported on five cohorts of patients from different studies. a=cohort 1; b=cohort 2; c=cohort 3; d=cohort 4; e=cohort 5

Mills 2016[a,b]: This study measured in the MID using two different anchors. a=walking anchor; b=knee health anchor

*Estimate is for improvement/deterioration

MC[a]=Mean change method proposed by Redelmeier

MC[b]=Mean change method proposed by Jaeshcke

Chapter 6: Discussion and Opportunities in Future MID Research

OVERVIEW

This thesis compiles a series of investigations focused on the synthesis, appraisal and application of anchor-based MID estimates in the context of enhancing interpretation of PROMs in clinical research – and thus, to guide clinical practice. This work is part of a wider emerging program of research to advance MID methodology, and represents an innovative contribution to a rapidly expanding field. This concluding chapter includes a discussion of key findings, limitations and opportunities for future directions.

IMPORTANT FINDINGS

This thesis commenced by responding to a critical issue that users of the MID are presently challenged with: the multiplicity of MIDs derived using diverse methodologies and no currently accepted standards for appraising the quality of these estimates. Failure to use trustworthy MIDs to provide interpretable estimates of treatment effects measured by PROMs may result in serious misinterpretations of the results of otherwise well-designed clinical trials and meta-analyses.

Based on this requirement to distinguish between more and less credible MID estimates, we developed a novel instrument to assess the credibility of anchor-based MID estimates – the instrument proved highly reliable. The widespread adoption and implementation of our instrument could have a twofold effect: 1) improved appraisal and encouraged use of credible MIDs in clinical trials, systematic reviews and clinical practice guidelines to inform evidence-based clinical and health policy decision making; and 2) the development of better-quality MID estimates.

To fully grasp the breadth of the MID literature and deepen our understanding of methodological issues and challenges involved in MID determinations, we conducted a systematic survey to develop a comprehensive compendium of published anchor-based MID estimates for all PROMs in the medical literature

from 1989 to 2015. We identified a large number of MID estimates available in the literature that investigators can use to inform the interpretation of PROMs across a wide variety of clinical areas; for many PROMs there are a number of MID estimates; and there are substantial deficiencies in the methodology of most MID assessments and a number of key improvements in anchor-based MID methodology remain possible.

Three insights that emerged could potentially lead to major improvements in the field. First, the profusion of varying terms representing the MID concept adds unnecessary and problematic complexity to users' task in comprehensively identifying relevant MIDs, requiring meticulous inspection of methodology to ensure estimates offered truly reflect the smallest possible difference that patients consider important. Second, a multitude of diverse methods for MID estimation exist; these methods yield varying estimates. Whether one or more represent better choices remains unresolved. Third, we found major deficiencies in reporting of fundamental characteristics of MID estimations that raised serious doubts and often precluded us from making judgements about aspects of MID credibility.

The application of MIDs for interpretation of PROM results requires users of the MID to conduct searches to identify reports of MIDs and when, as is often the case, the literature includes a number of candidate MIDs for a given PROM, choosing the most credible and applicable estimate is likely to prove difficult. Although the use of MIDs, as we demonstrated in Chapter 5 of this thesis, has proved invaluable for interpreting PROMs, the identification of MIDs in the medical literature represents an onerous task, particularly when the MID itself is not the primary focus of the investigation (e.g clinical trial or systematic review evaluating the effects of an intervention). Thus, the MID inventory will serve as a resource for users and developers of MIDs, simplifying their identification and utilization in clinical trials, systematic reviews and clinical practice guidelines.

The thesis ends with a real-word example demonstrating the practical value of MIDs in aiding a guideline panels' understanding of the importance of intervention effects for critically important outcomes (i.e. pain, physical function and quality of life) informing their recommendation. The search for credible anchor-based MIDs for PROMs evaluated in trials included in the linked systematic review comparing arthroscopic surgery to conservative management in patients with degenerative knee disease found, for the majority of the PROMs, that more than one study provided MID estimates. Moreover, many of the studies generated an MID at more than one duration of follow-up, used different anchors, and used various analytic methods, resulting in multiple estimates varying considerably in magnitude. Despite these challenges, we were able to provide a range of plausible trustworthy MIDs for key instruments that informed evidence presentation in the associated meta-analysis of treatment effects, and judgments by the guideline panel. The MIDs allowed the panel to weigh the magnitude of benefit against the harms of arthroscopy, and in doing so, the panel was confident making a strong recommendation against knee arthroscopy.

STRENGTHS AND LIMITATIONS

The strength of the work presented in this thesis is its innovation, which provides potential important advances in the field of PROs and MIDs. We have developed two important resources for users and developers of MIDs – a comprehensive inventory and an instrument to evaluate credibility – that will improve access and identification and appraisal of these estimates for enhancing the interpretation of PROMs and thereby facilitating healthcare decisions. Another strength is the scope and the breadth of rich information we have collected on available MIDs published in the literature, through which we have uncovered several opportunities for improving methodologic standards for anchor-based MID estimations and important gaps in research.

Beyond the limitations noted in the individual chapters, a limitation of this is thesis is that it does not address the development of knowledge translation (KT) and implementation strategies for the tools we have created. We have, however, identified key potential stakeholders who will likely directly benefit from our work: researchers, clinicians, patients, guideline developers and policy-makers. It will be incumbent upon our team leading this work to devise a KT plan in short order. Such a plan may include, in addition to publishing our work in medical journals, delivering workshops and presentations at international meetings. It will also be important to evaluate stakeholders' interest and perceived value of our work through evaluation of the overall reach, adoption and implementation of our tools.

This thesis focuses on the MID, which represents only one benchmark to enhance the interpretability of the PROMs. Indeed, there are other criteria proposed in the literature such as the patient acceptable symptom state (PASS) – the value below which the majority of patients consider themselves in an acceptable state of symptoms – and other responder definitions representing moderate or large magnitudes of effect, which may also aid in PROM interpretation. We strictly focused only anchor-based MID estimations, as opposed to distribution-based methods that are based on solely statistical properties of the sample and do not reflect patient importance. Despite these serious limitations, some researchers are using distribution-based methods to estimate MIDs and many have proposed using a combination of distribution- and anchor-based approaches. Our view is that, until there is convincing evidence that distribution-based methods are also valid for MID estimation, investigators should eschew such practices.

OPPORTUNITIES FOR FUTURE RESEARCH

This thesis demonstrates that there are several opportunities in the MID field to improve transparency and develop standards for research quality. There are also many questions in this area of research that still require exploration. For instance,

203

in developing the MID inventory, and in other related work[1], we found that the magnitude of MIDs for the same PROM often vary widely. Several other researcher groups have made similar observations stressing the importance of improved understanding of factors influencing the magnitude of MIDs to better inform decision making by interpreting PROMs with the most appropriate MID estimates[2-6]. For instance, researchers estimating MIDs for the WOMAC pain instrument using primary data from five studies in patients with hip or knee complaints, found large variation in estimates by the same analytic method (e.g. mean change) across studies and across different analytic methods (e.g. mean change versus ROC curve) within studies. The authors concluded that it was not possible to determine whether the observed variability in MID estimates was as a result of true differences in population and intervention characteristics, conceptual and methodological issues of MID analytic methods, or both. This variation suggests that the MID may depend on methodology and context, factors that one should therefore consider when appraising the credibility and applicability of published MIDs. Leveraging our large database of MIDs, we plan to conduct an empirical investigation exploring factors that may explain variability in MID estimates.

While exploring variability in the MID may help reveal methods that are more and less credible for MID estimation, ascertaining the implications of variability in MID estimated for trials and systematic reviews also requires consideration. We addressed this issue in the systematic review on treatment benefits and harms using MIDs that informed the BMJ Rapid Recommendation on knee arthroscopy for degenerative knee disease. The guideline authors found that results were robust to uncertainty in the MID. That is, estimates for the risk difference in the proportion of patients experiencing an important benefit with arthroscopy from sensitivity analyses using the upper and lower limit of the range of plausible MID estimates in sensitivity analyses yielded results similar to the primary analysis.

This is a single example and further exploration is required to systematically evaluate how variability in the MID may impact magnitude of treatment effects on PROMs. In the not too distant future, we plan on empirically testing this issue by conducting a re-analysis of meta-analyses assessing PROMs and examining how interpretation of pooled results may differ across alternative meta-analytic approaches established for PROMs[7-9], including those that do and do not rely on the MID.

A future investigation of high priority on our research agenda includes an in-depth evaluation of reporting issues in published anchor-based MID estimation articles. We will leverage our existing MID inventory to systematically evaluate the quality of reporting in these studies, with an emphasis on information necessary for distinguishing more from less trustworthy estimates. Currently, reporting guidance for MID estimation studies are unavailable. Thus, the proposed investigation and our prior work will inform the development of a formal consensus-based reporting checklist such as those offered by the EQUATOR Network. The development of a reporting standard for anchor-based MID estimation studies is likely to address, to an appreciable extent, many of the contemporary problems in the presentation of methods and results among MID studies.

Collectively, the proposed empirical investigations and the development of reporting standards will help address critical issues identified from this work that could potentially lead to major advancements in the field.

References

1. Devji T, Guyatt GH, Lytvyn L, et al. Application of minimal important differences in degenerative knee disease outcomes: a systematic review and case study to inform BMJ Rapid Recommendations. BMJ open. 2017;7(5):e015587.
2. Frahm Olsen M, Bjerre E, Hansen MD, Tendal B, Hilden J, Hrobjartsson A. Minimum clinically important differences in chronic pain vary considerably by baseline pain and methodological factors: systematic review of empirical studies. Journal of clinical epidemiology. 2018;101:87-106.e102.
3. Guyatt G, Schunemann H. How can quality of life researchers make their work more useful to health workers and their patients? Qual Life Res. 2007;16(7):1097-1105.
4. Terluin B, Eekhout I, Terwee CB. The anchor-based minimal important change, based on receiver operating characteristic analysis or predictive modeling, may need to be adjusted for the proportion of improved patients. Journal of clinical epidemiology. 2017;83:90-100.
5. Terluin B, Eekhout I, Terwee CB, de Vet HC. Minimal important change (MIC) based on a predictive modeling approach was more precise than MIC based on ROC analysis. Journal of clinical epidemiology. 2015;68(12):1388-1396.
6. Terwee CB, Roorda LD, Dekker J, et al. Mind the MIC: large variation among populations and methods. J Clin Epidemiol. 2010;63(5):524-534.
7. Guyatt GH, Thorlund K, Oxman AD, et al. GRADE guidelines: 13. Preparing summary of findings tables and evidence profiles-continuous outcomes. J Clin Epidemiol. 2013;66(2):173-183.
8. Johnston BC, Patrick DL, Thorlund K, et al. Patient-reported outcomes in meta-analyses-part 2: methods for improving interpretability for decision-makers. Health Qual Life Outcomes. 2013;11:211.
9. Thorlund K, Walter SD, Johnston BC, Furukawa TA, Guyatt GH. Pooling health-related quality of life outcomes in meta-analysis-a tutorial and review of methods for enhancing interpretability. Research synthesis methods. 2011;2(3):188-203.