Expanding the Definition of Clinical Differences: From Minimally Clinically Important Differences to Really Important Differences. Analyses in 8931 Patients with Rheumatoid Arthritis

FREDERICK WOLFE, KALEB MICHAUD, and VIBEKE STRAND

ABSTRACT. ***Objective.*** Minimally clinically important differences (MCID) have become an important way to interpret data of randomized clinical trials (RCT), but do not reflect the degree of improvement consistent with a "really important difference" (RID). To define RID, we compared mean and/or least desirable clinical states with best and/or most desirable states.

Methods. In total, 8931 patients with rheumatoid arthritis (RA) < 65 years of age completed the Health Assessment Questionnaire (HAQ) and Medical Outcomes Survey Short Form 36 Physical Component Score (PCS). Definitions of RID were based on values for HAQ and PCS corresponding with the best and worst category of the following conditions: disabled vs not disabled: joint replacement vs no joint replacement; ≤ poverty level vs > poverty level; very satisfied with health vs not; and independent in participation activities vs not.

Results. In contrast to published MCID values for the HAQ of ~0.22, RID was as high as 0.76 using objective reference conditions and 0.87 using the subjective measure of dependence vs independence. The HAQ score of independent RA patients was 0.38 (SD 0.45), and was 0.42 (SD 0.53) for those very satisfied with their health. The difference in HAQ scores between disabled and working patients was ~0.75. PCS differences were similarly increased.

Conclusion. RID values are 3 to 4 times greater than MCID values. Although MCID are meaningful statistics for RCT, the RID percentage achieved [(actual improvement/RID) × 100%] is a simple way to put the results of RCT in a broader perspective that gives an idea of how much additional treatment effect is needed. (J Rheumatol 2005;32:583–9)

Key Indexing Terms:
REALLY IMPORTANT DIFFERENCE RHEUMATOID ARTHRITIS
MINIMALLY CLINICALLY IMPORTANT DIFFERENCE

As the result of either therapeutic interventions or natural history, patients report changes in clinical status, measured by physical function, pain, and health related quality of life (HRQOL). When interpreting data from clinical care and randomized controlled trials (RCT) a key question is quantifying the amount of change that represents useful and clinically important improvement. A widely adopted approach for defining meaningful change is to identify the minimum clinically important difference (MCID), or the minimal detectable improvement perceptible to patients. Recent efforts to define MCID in RCT have suggested that improvements in patient reported outcomes of 33–36% over baseline (or 18% greater than placebo) are detectable as meaningful changes from baseline[1,2].

Assessment of physical function is perhaps the most important patient reported outcome in rheumatoid arthritis (RA), as it is a cumulative measure of the impact of disease and the best predictor of clinically important outcomes, including work disability[3-7], joint replacement[8], mortality[9], and medical costs[10]. The 2 most common assessments of functional status are the Health Assessment Questionnaire Disability Index (HAQ-DI)[11] and the generic Medical Outcomes Survey Short Form 36 (SF-36) measure of health related quality of life (HRQOL), including the Physical Component Summary score (PCS)[12,13]. Analyses defining MCID for HAQ-DI and SF-36 PCS in RCT have been based on statistical correlations with patient reported improvements in global disease activity and pain.

Interviewing 103 patients with RA, Redelmeier and Lorig determined that HAQ-DI scores needed to improve by –0.23 (95% CI 0.13, 0.23) units before respondents stopped rating themselves as about the same[14]. Wells, *et al* reported

From the Arthritis Research Center Foundation and the University of Kansas School of Medicine, Wichita, Kansas; the Center for Primary Care and Outcomes Research, and the Division of Immunology, Stanford University School of Medicine, Stanford, California, USA.

F. Wolfe, MD, National Data Bank for Rheumatic Diseases, University of Kansas School of Medicine; K. Michaud, MS, National Data Bank for Rheumatic Diseases; V. Strand, MD, Division of Immunology, Stanford University.

Address correspondence to Dr. F. Wolfe, National Data Bank for Rheumatic Diseases, Arthritis Research Center Foundation, 1035 N. Emporia, Suite 230, Wichita, KS 67214.
E-mail: fwolfe@arthritis-research.org

Accepted for publication November 9, 2004.

MCID to be 0.22 units in an RCT in active RA[2]. Zhao and Kosinski compared changes in HAQ-DI with patient assessments of global disease activity and pain in 2 RCT comparing COX-2 selective agents to traditional nonsteroidal anti-inflammatory drugs (NSAID) in active RA[15,16]. These analyses yielded good agreement (–0.24 to –0.22) with previously published values for MCID of –0.22[17]. Standard effect sizes from an RCT comparing leflunomide and methotrexate with placebo indicated both HAQ-DI and Modified Health Assessment Questionnaire (MHAQ) were sensitive to active treatment effects, HAQ-DI more than MHAQ[18].

Kosinski and Ware observed that mean changes in SF-36 domain scores corresponding to one level of improvement in patient reported pain or global assessment of disease activity ranged from 1.9 to 10.8 and 4.2 to 21.0, and 3.0 to 4.4 for SF-36 PCS and 2.2 to 4.7 for mental component summary (MCS) scores[15]. A variety of analyses have indicated that improvements of 5–10 points in SF-36 domains and 2.5–5 points in PCS and MCS summary scores represent MCID in RA, systemic lupus erythematosus, and osteoarthritis[1,2,15-25].

Analyzing changes in patient reported outcomes from RCT in the context of MCID helps to understand treatment data in the context of regular clinical practice. Nonetheless, MCID, rather than "important clinical improvement," raises a number of objections.

1. MCID represents a minimal clinically important (or detectable) change, which may be neither clinically meaningful nor useful, as detectable differences may simply be too small. And the importance of treatment associated differences may depend on baseline values as well as absolute change. For example, an improvement of 0.22 in HAQ-DI (defined as MCID) has a completely different meaning for baseline HAQ-DI scores ≥ 2 than when they are 0.25 to 1.0.
2. As MCID identifies a minimal detectable improvement rather than deterioration, it is not possible to interpret the magnitude of change patients perceive to be "important."
3. When applied to RCT, it is not always clear whether MCID should refer to absolute change from baseline or if one should also subtract the result of placebo or comparator treatment[26].
4. MCID does not offer clinicians an appropriate goal for improvement, based on patients' perceptions of realistic and desirable HAQ-DI or SF-36 PCS scores.
5. Therefore it is not clear how definitions of MCID should be used to interpret results from RCT or applied to clinical practice.

In contrast to minimal detectable improvements or MCID, in this report we investigate and define "really important differences" (RID) for HAQ-DI and SF-36 PCS in 8931 patients with RA. As there are no guidelines for defining "clinically important improvement," we based RID on objective outcomes of work disability, joint replacement surgery, and poverty as well as subjective reports of health satisfaction and functional independence. In addition, we report specific levels of HAQ-DI and SF-36 PCS associated with these outcomes — disabled versus not, total joint replacement versus no total joint replacement, poverty versus no poverty, satisfaction versus dissatisfaction with health, and dependence versus independence. Finally, we propose a percentage RID score that offers the amount of additional treatment effect required to attain these outcomes.

MATERIALS AND METHODS

Patient population. This study was performed using the National Data Bank for Rheumatic Diseases (NDB). The NDB is a research data bank where patients with rheumatic diseases complete detailed self-report questionnaires at 6 month intervals, as reported[27-30]. Eligible for this study were patients with RA who had completed at least one biannual survey for events between January 1, 1999, and December 31, 2001. A survey was selected at random for patients who had completed more than one. The resultant data set recorded 15,017 patients diagnosed with RA by their rheumatologist. Analyses were restricted to patients < 65 years of age to allow meaningful results regarding US Social Security disability benefits and work disability.

Demographic and disease status variables. NDB participants are asked to complete detailed semiannual 28 page questionnaires about all aspects of their illness. We record demographic variables including sex, age, ethnic origin, education level, current marital status, medical history, and total family income. Patient reported disease status and activity measures collected include HAQ[11,31], MHAQ[32], HAQ2[33], pain, global disease severity, and fatigue by visual analog scale[34], Rheumatoid Arthritis Disease Activity Index (RADAI), and SF-36[35-37]. Utilities are mapped from HAQ-DI values based on a regression model derived from simultaneous administration of the EuroQol instrument[38-40] and HAQ to 565 RA patients[41]. The EuroQol is a validated quality of life (QOL) scale that identifies 243 possible health states based on 5 questions concerning mobility, self-care, usual activity, pain/discomfort, and anxiety/depression; each item has 3 possible levels (1–3). "Utilities" or societal valuations have been placed on each state using time-tradeoff to determine valuations. Perfect health and death have utilities of 1 and 0, and states worse than death (< 0) are possible. Because it is short, the EuroQol can be particularly useful in surveys. Satisfaction with health is determined by asking, "How satisfied are you with your health now?" Choices are: very satisfied, somewhat satisfied, neither satisfied nor dissatisfied, somewhat dissatisfied, and very dissatisfied. Dependence on others is queried by: "How often do you have to depend on others for help?" Categories are: none of the time, a little of the time, some of the time, most of the time, and all of the time.

Receipt of US Social Security disability payments was used to define work disability in persons < 65 years of age. This definition excludes work disability in women and others not in the workforce and therefore not eligible for Social Security disability benefits, and excludes persons > 65 years of age who would work if they could. Therefore a second definition, self-reported work disability, was also utilized.

To determine poverty levels, US Health and Human Services (HHS) poverty guidelines for the 48 contiguous states for years 1998–2003 were applied[42]. The level selected for this study, 185% of the HHS poverty guideline, is commonly utilized as a measure of poverty, and to determine eligibility for the School Breakfast and Lunch programs. In this report we consider poverty as surrogate for income loss caused by RA. Level of educational attainment data in the general population were obtained from the US Census report of 2002[43].

Statistical methods. To determine differences based on disease severity characteristics listed in Table 2, Monte Carlo simulations with 1000 repetitions were performed. Analyses included adjustments for age and sex so that differences would be independent of these characteristics. Monte Carlo modeling was performed using Stata v. 8.2[44] and Clarify programs[45].

RESULTS

Clinical and demographic characteristics. Table 1 gives clinical and demographic variables in the 8931 study patients. Median disease duration was 9.7 years; 515 patients had ≤ 2 years of disease. Non-Hispanic whites made up 88.7% of the study sample; 28.2% had completed college; 8.5% had not completed high school. Mean HAQ-DI score was 1.07 (range 0–3), SF-36 PCS 32.4 (range 0–50), and mapped EuroQol utility 0.59 (range 0–1.0).

Clinically important differences. Figure 1 depicts HAQ scores based on health status (total joint replacement, poverty, reported disability, Social Security disability payments) and patient reported satisfaction and dependency. For comparison, the HAQ MCID difference is presented on the left. Data representing the dichotomous health status and patient reported states are presented in Table 2.

For "Satisfaction with health" HAQ-DI scores corresponding with very satisfied were 0.42 (0.53), somewhat satisfied: 0.83 (0.63), neither satisfied nor dissatisfied: 1.22 (0.64), somewhat dissatisfied: 1.35 (0.62), and very dissatisfied: 1.73 (0.59). Values at comparable points for the SF-36 PCS were 44.23 (8.18), 35.66 (8.89), 30.02 (8.55), 26.96 (7.68), and 22.78 (6.72).

RA patients receiving Social Security disability had lower mean HAQ-DI score differences of 0.76 (95% CI 0.72 to 0.76) units (Table 1, Figure 1). Using a less rigorous definition based on self-report of disability, a similar difference of 0.74 (95% CI 0.71 to 0.76) units was reported. Differences in mean HAQ-DI scores between individuals with and those without total joint replacements were less: 0.54 (95% CI 0.49 to 0.59) units, as for those with poverty: 0.57 (95% CI 0.52 to 0.61) units. When patient estimates of satisfaction and dependence on others were assessed, differences in HAQ-DI scores were greater. Patients very satisfied with their health differed from those who were not by 0.75 (95% CI 0.71 to 0.79) units. A greater difference was evident when comparing patients fully independent with those dependent on others: 0.87 (95% CI 0.83 to 0.91) units.

"Health" or "best" values for HAQ-DI. Patients who were fully satisfied with their health had a mean HAQ score of 0.42 and those who were fully independent had a mean HAQ score of 0.38 (Table 2).

Clinically important differences for SF-36 PCS. Similar differences in SF-36 PCS scores for patients working/not working were 9.56 (95% CI 9.08 to 10.04), Social Security disability 9.10 (95% CI 8.66 to 9.57), self-reported disability 6.47 (95% CI 5.92 to 7.04), and total joint replacement 5.85 (95% CI 5.09 to 6.60). Differences between patients very satisfied/not satisfied with their health and fully independent versus dependent on others were 13.72 (95% CI 13.12 to 14.29) and 11.84 (95% CI 11.17 to 12.58), respectively, with mean values of 44.23 and 41.45 (Table 2).

Relationship between differences in HAQ-DI and health utility scores. Reported differences in HAQ-DI and SF-36 PCS scores can also be expressed in terms of health utility scores. When adjusted for age and sex, differences in health utility scores for RA patients < 65 years of age, with disability and receiving Social Security benefits, were 0.23 (95% CI 023 to 0.24) and 0.22 (95% CI –21 to 0.23), respectively. Differences for poverty were 0.19 (95% CI 0.18 to 0.21) and total joint replacement 0.12 (95% CI 0.10 to 0.14). As with HAQ-DI and SF-36 PCS, differences were greatest between patients very satisfied with their health and fully independent versus not satisfied and dependent on others: 0.27 (95% CI 0.25 to 0.28) and 0.24 (95% CI 0.22 to 0.26), respectively. In comparison, a difference in HAQ-DI scores of 0.22 or MCID corresponded with a difference in health utilities of 0.06 (95% CI 0.06 to 0.06).

Table 1. Clinical and demographic variables for 8931 patients with RA.

Variable	Mean	SD	Median
Age, yrs	51.6	9.4	53.1
Sex, % male	20.3		
Education, years	13.6	2.3	13.0
Education category, %			
0–8	1.7		
8–11	6.8		
12	35.1		
13–15	28.2		
≥ 16	28.2		
Ethnic origin, %			
White	88.7		
Black	4.8		
Asian-American	1.2		
Indian or Am. Indian	1.3		
Mexican or Mexican American	3.1		
Puerto Rican	0.4		
Other	0.5		
Total income, US dollars	50,000	2930	45,000
Lifetime comorbidity score (0–11)	2.0	1.7	2.0
Disease duration, years	12.1	9.5	9.7
HAQ (0–3)	1.07	0.7	1.1
HAQ2 (0–3)	1.01	0.6	1.0
MHAQ (0–3)	0.54	0.5	0.5
RADAI score (0–10)	3.8	2.2	3.7
Pain (0–10)	4.0	2.8	3.5
Global severity (0–10)	3.5	2.6	3.0
Fatigue (0–10)	4.7	3.0	5.0
Depression (0–10)	2.7	1.8	2.3
Physical component score, SF-36	32.4	10.5	31.6
Mental component score, SF-36	42.6	14.4	45.3
Mapped EuroQol utility (0–1)	0.59	0.24	0.65
Very satisfied with health, %	14.3		
Not dependent on others, %	23.6		

DISCUSSION

To persons not familiar with HAQ-DI and SF-36 PCS scores, data from RCT and clinical practice can be confusing and difficult to interpret. For example, what is a good HAQ-DI or SF-36 PCS score, or what is a clinically important improvement? Some answers to these questions are pro-

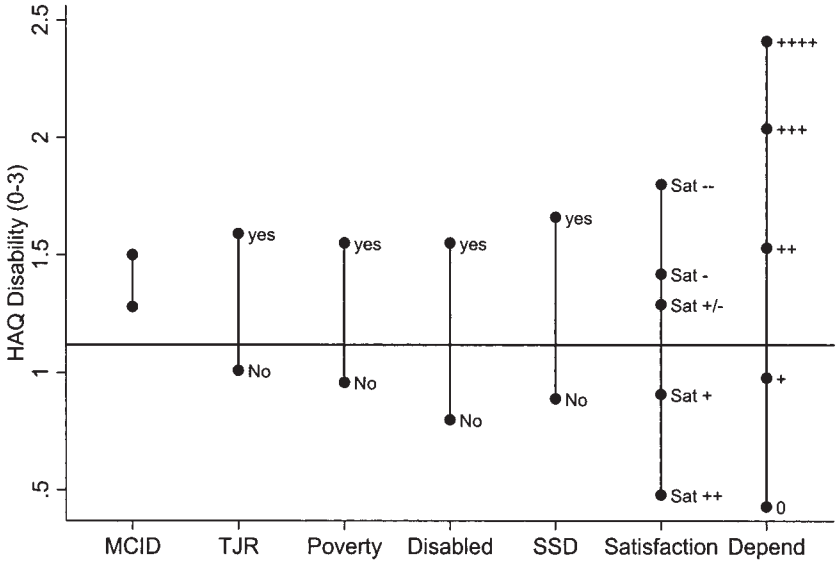

Figure 1. Mean Health Assessment Questionnaire (HAQ) scores for key RA outcomes. The length of the line represents the difference in HAQ score between outcome categories. The MCID difference of 0.22 from the literature was placed arbitrarily at 1.50 and 1.28 units, 1.50 representing a value that could be found in an RCT. MCID: Minimal clinically important difference; TJR: total joint replacement; Poverty: 180% US poverty level; SSD: Social Security disability award; Satisfaction: satisfaction with health [categories are very satisfied (Sat++), somewhat satisfied (Sat+), neither satisfied nor dissatisfied (Sat+/–), somewhat dissatisfied (Sat–), very dissatisfied (Sat– –)]; Depend: Dependence on others (categories: none of the time, a little of the time, some of the time, most of the time, all of the time).

Table 2. Really important differences in HAQ and SF-36. Results are adjusted for age and sex, and refer to patients < 65 years of age.

Status	Difference*	State	
	Mean (95% CI)	Mean (SD)	Mean (SD)
Work disabled		Not disabled	Disabled
HAQ (0–3)	0.74 (0.71, 0.76)	0.80 (0.63)	1.55 (0.63)
Physical component score, SF-36	9.1 (8.6, 9.57)	35.61 (9.96)	26.32 (8.57)
Social Security disability		Not disabled	Disabled
HAQ (0–3)	0.76 (0.72, 0.79)	0.89 (0.67)	1.66 (0.56)
Physical component score, SF-36	9.56 (9.08, 10.04)	34.56 (10.19)	24.80 (7.57)
Total joint replacement (TJR)		No TJR	TJR
HAQ (0–3)	0.54 (0.49, 0.59)	1.01 (0.71)	1.59 (0.65)
Physical component score, SF-36	5.85 (5.09, 6.60)	32.95 (10.49)	28.84 (8.64)
Poverty		No poverty	Poverty
HAQ (0–3)	0.57 (0.52, 0.61)	0.96 (0.69)	1.55 (0.66)
Physical component score, SF-36	6.47 (5.92, 7.04)	33.46 (10.40)	26.72 (8.80)
Satisfied with health		Very satisfied	Not very satisfied
HAQ (0–3)	0.75 (0.71, 0.79)	0.42 (0.53)	1.17 (0.70)
Physical component score, SF-36	13.72 (13.12, 14.29)	44.23 (8.18)	30.48 (9.52)
Depend on others for help		Fully independent	Not independent
HAQ (0–3)	0.87 (0.83, 0.91)	0.38 (0.45)	1.27 (0.64)
Physical component score, SF-36	11.84 (11.17, 12.58)	41.45 (8.66)	29.62 (9.13)

* Based on Monte Carlo simulation studies with 1000 repetitions.

vided in Table 2. Patients with RA who are independent and very satisfied with their health report low HAQ-DI scores ≤ 0.375. Patients who have had joint replacements, are work disabled, or are living at a poverty level report high HAQ-DI scores ≥ 1.5. Similarly, patients with good outcomes report SF-36 PCS scores between 41.5 and 44.2, whereas those with joint replacements, with disability, and who live at poverty level have SF-36 PCS scores between

26.3 and 28.8. Figure 2 (HAQ values at categories of health satisfaction) and Figure 3 (PCS values at categories of health satisfaction) provide intermediate values for HAQ-DI and SF-36 PCS scores that offer additional perspective. Based on categorical ratings to the questions above, if a HAQ-DI score ≤ 0.42 is "very good," then scores of ~0.83 may be interpreted as "good," ~1.22 and ~1.34 as "poor," and ≥ 1.73 as "very poor." Similar divisions for the SF-36 PCS are presented in Figure 3 and the Results section.

More simply, one might use the 2 extreme HAQ (0.375 and 1.50) and PCS values (43 and 27) as markers for best and worst categories and divide the values between them into one or 2 additional cutpoints, as both the HAQ and PCS are linear at these points in their scale. The largest HAQ and SF-36 RID were related to work disability, considering objective outcomes. However, even greater differences were noted in health satisfaction and independence, indicating that subjective impressions by patients can provide further insight.

As expected, RID are considerably greater than MCID.

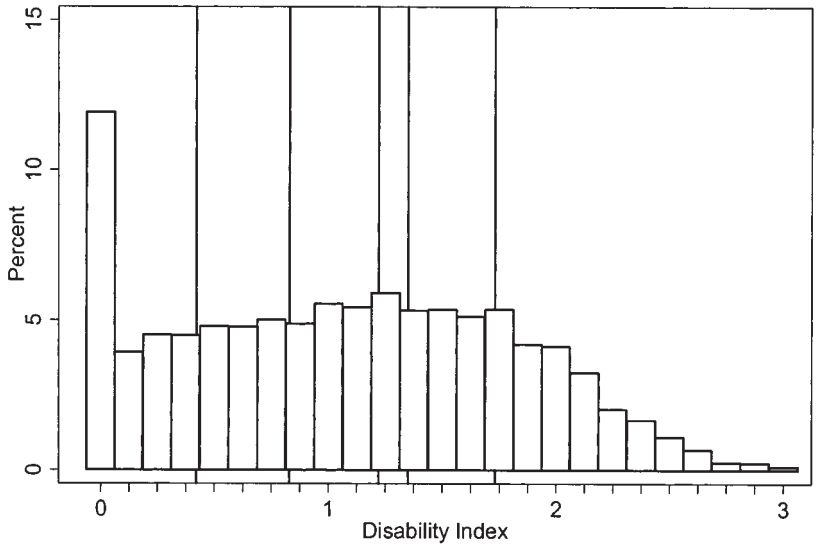

Figure 2. Distribution of HAQ values for RA patients < 65 years of age. Vertical lines correspond to HAQ values for categories of satisfaction with health: very satisfied, somewhat satisfied, neither satisfied nor dissatisfied, somewhat dissatisfied, and very dissatisfied. Mean (SD) values were 0.42 (0.53), 0.83 (0.63), 1.22 (0.64), 1.35 (0.62), and 1.73 (0.59), respectively. Marks on the x-axis occur every 0.125 units.

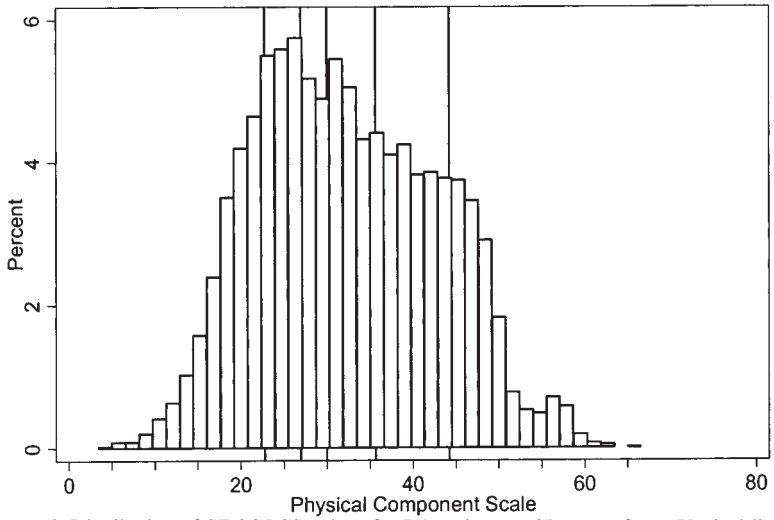

Figure 3. Distribution of SF-36 PCS values for RA patients < 65 years of age. Vertical lines correspond to HAQ values for categories of satisfaction with health: very satisfied, somewhat satisfied, neither satisfied nor dissatisfied, somewhat dissatisfied, and very dissatisfied. Mean (SD) values were 44.23 (8.18), 35.66 (8.89), 30.02 (8.55), 26.96 (7.68), and 22.78 (6.72), respectively.

Definitions of RID complement MCID by placing the concept of MCID in a broader perspective. Using a health utility score as a common metric, improvements corresponding with MCID result in small differences of 0.06, whereas RID differences based on satisfaction with health, independence, and no work disability are as great as 0.27, 0.26, and 0.23, respectively. Just as the MCID constitutes a necessary minimum, the RID represents a clinically important goal.

Whereas MCID helps to interpret mean changes across treatment groups in a clinical trial, RID is a useful technique to measure improvement on an individual basis, both in RCT and in clinical practice. One approach would be to measure the RID percentage achieved [(actual improvement/RID) × 100]. For example, if a patient reported improvement in HAQ-DI score by 0.22, or MCID, his RID percentage would be 29.3% using 0.75 as RID (based on satisfaction with health) and 25.3% using 0.87 as RID (based on independence). An RID percentage, therefore, offers a simple way to put the results of clinical trials into a broader perspective.

REFERENCES

1. Goldsmith CH, Boers M, Bombardier C, Tugwell P. Criteria for clinically important changes in outcomes: development, scoring and evaluation of rheumatoid arthritis patient and trial profiles. OMERACT Committee. J Rheumatol 1993;20:561-5.
2. Wells GA, Tugwell P, Kraag GR, Baker PR, Groh J, Redelmeier DA. Minimum important difference between patients with rheumatoid arthritis: the patient's perspective. J Rheumatol 1993;20:557-60.
3. Wolfe F, Hawley DJ. The longterm outcomes of rheumatoid arthritis: work disability: a prospective 18 year study of 823 patients. J Rheumatol 1998;25:2108-17.
4. Sokka T, Pincus T. Markers for work disability in rheumatoid arthritis. J Rheumatol 2001;28:1718-22.
5. Young A, Dixey J, Kulinskaya E, et al. Which patients stop working because of rheumatoid arthritis? Results of five years' follow up in 732 patients from the Early RA Study (ERAS). Ann Rheum Dis 2002;61:335-40.
6. Doeglas D, Suurmeijer T, Krol B, Sanderman R, van Leeuwen M, van Rijswijk M. Work disability in early rheumatoid arthritis. Ann Rheum Dis 1995;54:455-60.
7. Makisara GL, Makisara P. Prognosis of functional capacity and work capacity in rheumatoid arthritis. Clin Rheumatol 1982;1:117-25.
8. Wolfe F, Zwillich SH. The long-term outcomes of rheumatoid arthritis: A 23-year prospective, longitudinal study of total joint replacement and its predictors in 1,600 patients with rheumatoid arthritis. Arthritis Rheum 1998;41:1072-82.
9. Wolfe F, Michaud K, Gefeller O, Choi HK. Predicting mortality in patients with rheumatoid arthritis. Arthritis Rheum 2003;48:1530-42.
10. Michaud K, Messer J, Choi HK, Wolfe F. Direct medical costs and their predictors in persons with rheumatoid arthritis: a 3 year study of 7,527 patients. Arthritis Rheum 2003;48:2750-62.
11. Fries JF, Spitz PW, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. Arthritis Rheum 1980;23:137-45.
12. Keller SD, Majkut TC, Kosinski M, Ware JE Jr. Monitoring health outcomes among patients with arthritis using the SF-36 Health Survey: overview. Med Care 1999;37:MS1-9.
13. Ware JE, Sherbourne CD. The MOS 36-Item Short-Form Health Survey (SF-36). 1. Conceptual framework and item selection. Med Care 1992;30:473-83.
14. Redelmeier DA, Lorig K. Assessing the clinical importance of symptomatic improvements — an illustration in rheumatology. Arch Intern Med 1993;153:1337-42.
15. Kosinski M, Zhao SZ, Dedhiya S, Osterhaus JT, Ware JE. Determining minimally important changes in generic and disease-specific health-related quality of life questionnaires in clinical trials of rheumatoid arthritis. Arthritis Rheum 2000;43:1478-87.
16. Zhao SZ, Fiechtner JI, Tindall EA, et al. Evaluation of health-related quality of life of rheumatoid arthritis patients treated with celecoxib. Arthritis Care Res 2000;13:112-21.
17. Strand V, Cannon G, Cohen S, Ware J. Correlation of HAQ with SF-36: comparision of leflunomide to methotrexate inpatients with active RA [abstract]. Arthritis Rheum 2001;44 Suppl:S187.
18. Tugwell P, Wells G, Strand V, et al. Clinical improvement as reflected in measures of function and health-related quality of life following treatment with leflunomide compared with methotrexate in patients with rheumatoid arthritis: sensitivity and relative efficiency to detect a treatment effect in a twelve-month, placebo-controlled trial. Leflunomide Rheumatoid Arthritis Investigators Group. Arthritis Rheum 2000;43:506-14. Erratum: Arthritis Rheum 2000;43:1345.
19. Zhao SZ, McMillen JI, Markenson JA, et al. Evaluation of the functional status aspects of health-related quality of life of patients with osteoarthritis treated with celecoxib. Pharmacotherapy 1999;19:1269-78.
20. Beaton DE, Bombardier C, Katz JN, et al. Looking for important change/differences in studies of responsiveness. OMERACT MCID Working Group. J Rheumatol 2001;28:400-5.
21. Angst F, Aeschlimann A, Stucki G. Smallest detectable and minimal clinically important differences of rehabilitation intervention with their implications for required sample sizes using WOMAC and SF-36 quality of life measurement instruments in patients with osteoarthritis of the lower extremities. Arthritis Rheum 2001;45:384-91.
22. Ehrich EW, Davies GM, Watson DJ, Bolognese JA, Seidenberg BC, Bellamy N. Minimal perceptible clinical improvement with the Western Ontario and McMaster Universities osteoarthritis index questionnaire and global assessments in patients with osteoarthritis. J Rheumatol 2000;27:2635-41.
23. Angst F, Aeschlimann A, Stucki G. Smallest detectable and minimal clinically important differences of rehabilitation intervention with their implications for required sample sizes using WOMAC and SF-36 quality of life measurement instruments in patients with osteoarthritis of the lower extremities. Arthritis Rheum 2001;45:384-91.
24. Samsa G, Edelman D, Rothman ML, Williams GR, 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;15:141-55.
25. Strand V, Kelman A. Outcome measures in osteoarthritis: randomized controlled trials. Curr Rheumatol Rep 2004;6:20-30.
26. Strand V, Pincus T. The Health Assessment Questionnaire (HAQ) provides a single effective measure to discriminate active from placebo treatment in randomized controlled trials in rheumatoid arthritis [abstract]. Arthritis Rheum 2004;48 Suppl:3656.
27. Wolfe F, Anderson J, Burke TA, Arguelles LM, Pettitt D. Gastroprotective therapy and risk of gastrointestinal ulcers: risk reduction by COX-2 therapy. J Rheumatol 2002;29:467-73.
28. Wolfe F, Flowers N, Burke TA, Arguelles LM, Pettitt D. Increase in lifetime adverse drug reactions, service utilization, and disease severity among patients who will start COX-2 specific inhibitors:

quantitative assessment of channeling bias and confounding by indication in 6689 patients with rheumatoid arthritis and osteoarthritis. J Rheumatol 2002;29:1015-22.
29. Wolfe F, Flowers N, Anderson J. The National Rheumatic Disease Data Bank: case mix and severity characteristics of patients in rheumatological practice [abstract]. Arthritis Rheum 1998;41 Suppl:S132.
30. Wolfe F, Michaud K. Lymphoma in rheumatoid arthritis: the effect of methotrexate and anti-tumor necrosis factor therapy in 18,572 patients. Arthritis Rheum 2004;50:1740-51.
31. Fries JF, Spitz PW, Young DY. The dimensions of health outcomes: the Health Assessment Questionnaire, disability and pain scales. J Rheumatol 1982;9:789-93.
32. Pincus T, Summey JA, Soraci SA Jr, Wallston KA, Hummon NP. Assessment of patient satisfaction in activities of daily living using a modified Stanford Health Assessment Questionnaire. Arthritis Rheum 1983;26:1346-53.
33. Wolfe F, Michaud K, Pincus T. HAQ-II: Development and validation of a revised version of the health assessment questionnaire (HAQ). Arthritis Rheum 2004;50:3296-305.
34. Wolfe F, Hawley DJ, Wilson K. The prevalence and meaning of fatigue in rheumatic disease. J Rheumatol 1996;23:1407-17.
35. Stucki G, Liang MH, Stucki S, Bruhlmann P, Michel BA. A self-administered rheumatoid arthritis disease activity index (RADAI) for epidemiologic research. Psychometric properties and correlation with parameters of disease activity. Arthritis Rheum 1995;38:795-8.
36. Fransen J, Hauselmann H, Michel BA, Caravatti M, Stucki G. Responsiveness of the self-assessed Rheumatoid Arthritis Disease Activity Index to a flare of disease activity. Arthritis Rheum 2001;44:53-60.
37. Fransen J, Langenegger T, Michel BA, Stucki G. Feasibility and validity of the RADAI, a self-administered rheumatoid arthritis disease activity index. Rheumatology Oxford 2000;39:321-7.
38. Fransen M, Edmonds J. Reliability and validity of the EuroQol in patients with osteoarthritis of the knee. Rheumatology Oxford 1999;38:807-13.
39. Kobelt G, Eberhardt K, Jonsson L, Jonsson B. Economic consequences of the progression of rheumatoid arthritis in Sweden. Arthritis Rheum 1999;42:347-56.
40. EuroQol Group. EuroQol — a new facility for the measurement of health-related quality of life. Health Policy 1990;16:199-208.
41. Choi HK, Michaud K, Wolfe F. Utility measures in rheumatic disorders [abstract]. Arthritis Rheum 2002;46 Suppl:S76.
42. The 1998-2003 HHS poverty guidelines. Washington, DC: US Department of Health and Human Services; 2003. Available at: http://aspe.hhs.gov/poverty/98poverty.htm. Accessed January 13, 2005.
43. Table 5a. Educational attainment of civilians 16 years and over, by labor force status, age, sex, race, and Hispanic origin: March 2002. US Census 2003. Available at: http://www.census.gov/population/socdemo/education/ppl-169/tab05a.pdf. Accessed January 13, 2005.
44. Stata Corp. Stata statistical software: release 8.2. College Station, TX: Stata Corp.; 2003.
45. Tomz M, Wittenberg J, King G. CLARIFY: Software for interpreting and presenting statistical results; 1998. Available from: http://gking.harvard.edu/clarify/docs/clarify.html. Accessed January 13, 2005.