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
MCID to be 0.22 units in an RCT in active RA. 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. These analyses yielded good agreement (−0.24 to −0.22) with previously published values for MCID of −0.25. 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.

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. 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.

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.

4. MCID does not offer clinicians an appropriate goal for improvement, based on patient variations 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. Eligible for this study were patients with RA who had completed at least one bimannual 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 HAQ1, MHAQ3, HAQ2, pain, global disease severity, and fatigue by visual analog scale. Rheumatoid Arthritis Disease Activity Index (RADAI), and SF-36 PCS and MCS utilities are mapped from HAQ-DI values based on a regression model derived from simultaneous administration of the EuroQol instrument and HAQ to 565 RA patients. 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. 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.
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: 1.73 (0.59). Values at comparable points for the SF-36 were 35.66 (8.89) and 30.02 (8.55), respectively. 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 0.23 to 0.24) and 0.22 (95% CI –0.21 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).

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-
Provided 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

---

**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.

<table>
<thead>
<tr>
<th>Status</th>
<th>Difference* Mean (95% CI)</th>
<th>State</th>
<th>Mean (SD)</th>
<th>State</th>
<th>Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Work disabled</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HAQ (0–3)</td>
<td>0.74 (0.71, 0.76)</td>
<td>Not disabled</td>
<td>0.80 (0.63)</td>
<td>Disabled</td>
<td>1.55 (0.63)</td>
</tr>
<tr>
<td>Physical component score, SF-36</td>
<td>9.1 (8.6, 9.57)</td>
<td></td>
<td>35.61 (9.96)</td>
<td></td>
<td>26.32 (8.57)</td>
</tr>
<tr>
<td>Social Security disability</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HAQ (0–3)</td>
<td>0.76 (0.72, 0.79)</td>
<td>Not disabled</td>
<td>0.89 (0.67)</td>
<td>Disabled</td>
<td>1.66 (0.56)</td>
</tr>
<tr>
<td>Physical component score, SF-36</td>
<td>9.56 (9.08, 10.04)</td>
<td></td>
<td>34.56 (10.19)</td>
<td></td>
<td>24.80 (7.57)</td>
</tr>
<tr>
<td>Total joint replacement (TJR)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HAQ (0–3)</td>
<td>0.54 (0.49, 0.59)</td>
<td>Not TJR</td>
<td>1.01 (0.71)</td>
<td>TJR</td>
<td>1.59 (0.65)</td>
</tr>
<tr>
<td>Physical component score, SF-36</td>
<td>5.85 (5.09, 6.60)</td>
<td></td>
<td>32.95 (10.49)</td>
<td></td>
<td>28.84 (8.64)</td>
</tr>
<tr>
<td>Poverty</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HAQ (0–3)</td>
<td>0.57 (0.52, 0.61)</td>
<td>Not poverty</td>
<td>0.96 (0.69)</td>
<td>Poverty</td>
<td>1.55 (0.66)</td>
</tr>
<tr>
<td>Physical component score, SF-36</td>
<td>6.47 (5.92, 7.04)</td>
<td></td>
<td>33.46 (10.40)</td>
<td></td>
<td>26.72 (8.80)</td>
</tr>
<tr>
<td>Satisfied with health</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HAQ (0–3)</td>
<td>0.75 (0.71, 0.79)</td>
<td>Very satisfied</td>
<td>0.42 (0.53)</td>
<td>Not very satisfied</td>
<td>1.17 (0.70)</td>
</tr>
<tr>
<td>Physical component score, SF-36</td>
<td>13.72 (13.12, 14.29)</td>
<td></td>
<td>44.23 (8.18)</td>
<td></td>
<td>30.48 (9.52)</td>
</tr>
<tr>
<td>Depend on others for help</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HAQ (0–3)</td>
<td>0.87 (0.83, 0.91)</td>
<td>Fully independent</td>
<td>0.38 (0.45)</td>
<td>Not independent</td>
<td>1.27 (0.64)</td>
</tr>
<tr>
<td>Physical component score, SF-36</td>
<td>11.84 (11.17, 12.58)</td>
<td></td>
<td>41.45 (8.66)</td>
<td></td>
<td>29.62 (9.13)</td>
</tr>
</tbody>
</table>

* Based on Monte Carlo simulation studies with 1000 repetitions.
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.
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


44. Stata Corp. Stata statistical software: release 8.2. College Station, TX: Stata Corp.; 2003.