The Clinically Important Difference on the Unified Parkinson’s Disease Rating Scale

Lisa M. Shulman, MD; Ann L. Gruber-Baldini, PhD; Karen E. Anderson, MD; Paul S. Fishman, MD, PhD; Stephen G. Reich, MD; William J. Weiner, MD

Objective: To determine the estimates of minimal, moderate, and large clinically important differences (CIDs) for the Unified Parkinson’s Disease Rating Scale (UPDRS).

Design: Cross-sectional analysis of the CIDs for UPDRS total and motor scores was performed on patients with Parkinson disease (PD) using distribution- and anchor-based approaches based on the following 3 external standards: disability (10% on the Schwab and England Activities of Daily Living Scale), disease stage (1 stage on the Hoehn and Yahr Scale), and quality of life (1 SD on the 12-Item Short Form Health Survey).

Setting: University of Maryland Parkinson Disease and Movement Disorders Center.

Patients: Six hundred fifty-three patients with PD.

Results: A minimal CID was 2.3 to 2.7 points on the UPDRS motor score and 4.1 to 4.5 on the UPDRS total score. A moderate CID was 4.5 to 6.7 points on the UPDRS motor score and 8.5 to 10.3 on the UPDRS total score. A large CID was 10.7 to 10.8 points on the UPDRS motor score and 16.4 to 17.8 on the UPDRS total score.

Conclusions: Concordance among multiple approaches of analysis based on subjective and objective data show that reasonable estimates for the CID on the UPDRS motor score are 2.5 points for minimal, 5.2 for moderate, and 10.8 for large CIDs. Estimates for the UPDRS total score are 4.3 points for minimal, 9.1 for moderate, and 17.1 for large CIDs. These estimates will assist in determining clinically meaningful changes in PD progression and response to therapeutic interventions.

Arch Neurol. 2010;67(1):64-70

A clinically important difference (CID) is the amount of change on a measure that patients can recognize and value.1 Growing interest in CIDs stems from a greater emphasis on evidence-based and patient-centered medicine.2 Large randomized clinical trials frequently show significant differences on outcome measures that are so small that clinicians are unsure how to apply them to clinical decision making.3 The Movement Disorder Society Task Force on Rating Scales for Parkinson’s Disease highlighted the importance of identifying thresholds on the Unified Parkinson’s Disease Rating Scale (UPDRS) that represent clinically relevant differences.4,5 The US Food and Drug Administration also described the need to define minimally important differences on patient-reported outcome measures used to support the labeling claims of medical products.6

The 2 key methods of CID assessment are the distribution- and anchor-based approaches.7,8 The distribution-based approach relies on the empirical distribution of a measure in a population and the derived effect size. The anchor-based approach uses a familiar and relevant external standard to determine the corresponding magnitude of change. This study uses both of these approaches and relies on 3 different external standards (anchors) to assess the CID on the UPDRS total scale and its motor subscale.

The minimal clinically important change on the UPDRS was previously determined based on data from 2 clinical trials of dopamine agonist monotherapy in early Parkinson disease (PD).9 Study limitations included the inability to generalize the results to more advanced PD and the reliance on a clinician-based measure (Clinical Global Impression of Improvement) as the anchor for assessing clinical relevance. Assessments by clinicians do not always match patient evaluations because of limitations in awareness of the patient experience.6,10 The belief that clinically relevant differences in health

Author Affiliations:
Departments of Neurology (Drs Shulman, Anderson, Fishman, Reich, and Weiner), Epidemiology and Preventive Medicine (Dr Gruber-Baldini), and Psychiatry (Dr Anderson), University of Maryland School of Medicine, Baltimore.

©2010 American Medical Association. All rights reserved.

Downloaded From: http://archneur.jamanetwork.com/ on 04/18/2017
should be defined by patients is fundamental to patient-centered medicine.

The primary objective of this study was to determine the CID for the UPDRS by using multiple methods of assessment and a large patient sample representing all stages of PD. The goal was to create estimates of minimal, moderate, and large CIDs by looking for concordance of the results from multiple approaches of CID analysis.

**METHODS**

The sample consists of patients diagnosed as having PD by a movement disorder specialist (L.M.S., P.S.F., S.G.R., or W.J.W.) at the University of Maryland Parkinson Disease and Movement Disorders Center who underwent assessment during routine office visits from April 1, 2003, through August 31, 2006. The criteria for the diagnosis of PD were asymmetrical onset of at least 2 of the following 3 cardinal signs: resting tremor, rigidity, and bradykinesia, with no atypical signs or exposure to dopamine-blocking drugs. Patients attending the movement disorders center are routinely asked to enroll in the University of Maryland Quality of Life and Function Study. During the study period, 86% of patients with PD agreed to participate and signed an informed consent form approved by the University of Maryland institutional review board. The treating neurologist completed the UPDRS, staging with the Hoehn and Yahr Scale (HY), the Schwab and England Activities of Daily Living Scale (SE Scale), and the Mini-Mental State Examination for all subjects. Patients with a Mini-Mental State Examination score of less than 26 required the assistance of a caregiver for consent and questionnaire completion. Patients completed the 12-Item Short Form Health Survey; UPDRS, Unified Parkinson’s Disease Rating Scale.

In anchor-based methods, the measures chosen as anchors should be familiar to clinicians in the field, relevant, interpretable, and significantly correlated with the instrument being explored. Three measures were used as anchors in this study: (1) the SF-12, (2) the SE Scale, and (3) the HY stages. Pearson correlations were performed, showing that the UPDRS total and motor scores have moderate to large correlations with the SE Scale ($r = -0.64$ to $-0.78$), HY stages ($r = 0.70$ to $0.75$), and the SF-12 ($r = -0.44$ to $-0.52$), mental health ($r = -0.35$ to $-0.45$) (for all, $P < .001$). Previously accepted thresholds for the CID have been published for the 36-Item Short Form Health Status Survey (SF-36) and the SF-12, but thresholds have not been defined for the SE Scale or the HY stages.

**DETERMINING THE CID WITH DISTRIBUTION- AND ANCHOR-BASED ANALYSES**

There are many accepted methods of assessing CID; because all methods have strengths and weaknesses, it is preferable to rely on several methods. There is also no single response or precise threshold for the CID of a measure; instead it is best to represent the CID of a measure as a range (eg, small, moderate, or large). Therefore, in this cross-sectional study we used a combination of distribution- and anchor-based approaches, and we designated predetermined cut points of small, moderate, and large CIDs for the UPDRS.

For the distribution-based approach, means and standard deviations were derived from the current sample of data, and effect sizes were calculated relative to 1 SD. The most common approach to the analysis of the distribution-based CID in the literature relies on the Cohen effect size, in which an effect size of 0.2 (0.2 of an SD) is small, 0.5 is moderate, and 0.8 is large. In anchor-based methods, the measures chosen as anchors should be familiar to clinicians in the field, relevant, interpretable, and significantly correlated with the instrument being explored. Three measures were used as anchors in this study: (1) the SF-12, (2) the SE Scale, and (3) the HY stages. Pearson correlations were performed, showing that the UPDRS total and motor scores have moderate to large correlations with the SE Scale ($r = -0.64$ to $-0.78$), HY stages ($r = 0.70$ to $0.75$), and the SF-12 ($r = -0.44$ to $-0.52$), mental health ($r = -0.35$ to $-0.45$) (for all, $P < .001$). Previously accepted thresholds for the CID have been published for the 36-Item Short Form Health Status Survey (SF-36) and the SF-12, but thresholds have not been defined for the SE Scale or the HY stages.

**CUT POINTS FOR THE SF-36 AND SF-12**

The SF-36 and SF-12 have 2 summary scores—PH and MH—that yield $t$ scores based on a US normative population in which the average score is the 50th percentile and 10 units is 1 SD. An analysis of effect sizes for the SF-36 was performed by Samsa et al for about 25 medical conditions. Conforming to clinical intuition, conditions such as congestive heart failure or emphysema were associated with large effect sizes on the SF-36, conditions such as arthritis had moderate effect sizes, and conditions such as hypertension had small effect sizes. Based on the published literature, the small CID for the SF-36 or the SF-12 is in the range of 3 to 5 points, whereas the moderate CID is 9 to 10 points. These ranges for small and moderate CID were used in our study.

**CUT POINTS FOR THE SE SCALE AND HY STAGES**

In the absence of previous analysis of thresholds for effect sizes on the SE Scale or the HY stages, we used a combination of clinical judgment and analysis of each scale’s distribution (based on the standard deviation). Specifically, we made a predetermined judgment of small, medium, or large CID on the SE Scale and the HY stages based on our clinical experience. Then we analyzed the SE Scale and HY stage distributions in our sample to assess whether they conformed to our clinical impressions. On the SE Scale, clinicians assign scores based on descriptors that coincide with 10% increments on the scale; therefore, a 10% change (10 points on the SE Scale) is clinically relevant. Furthermore, the standard deviation on the SE Scale for our sample was 18.7 (Table 1). Therefore, our criterion for a 10% change is 0.53 SD (10 per 18.7) or approximately half of an SD for every 10% change: a moderate CID based on the Cohen effect size. On the HY stages, clinicians assign the 5 stages based on the clinical descriptions at each stage. Because PD is a gradu-
The mean UPDRS total score (subscales I, II, and III) was 41.0 (20.5). Based on the SE Scale ratings, the mean UPDRS total score ranged from a low of 19.9 (9.5) for subjects reporting no disability (SE Scale, 100% [completely independent]) to a high of 60.0 (7.1) for subjects rated as totally dependent (SE Scale, 10% [bedridden]) (Table 2). Based on the HY stages, the mean UPDRS total score ranged from a low of 11.2 (4.9) for subjects with unilateral parkinsonism (stage 1) to a high of 54.4 (11.4) for subjects assessed as wheelchair bound or bedridden (stage 5) (Table 3). Regression analysis showed the average change on the UPDRS total score to be 4.5 points for a 10% change on the SE Scale and 10.8 points for a 1-stage change on the HY stages. On the SF-12, 1 unit on the PH or the MH summary score was equivalent to a change of 0.47 points on the UPDRS motor score. Therefore, a difference of 1 SD (defined as 10 units) was 4.7 points on the UPDRS motor score and 8.6 points for the large CID (Table 4). The distribution-based analysis showed that the minimal CID was 2.7 points, the moderate CID was 6.7, and the large CID was 10.7 (Table 4). Based on a combination of the anchor- and distribution-based analyses (averaging across the results), 2.5 points is an appropriate estimate for the minimal CID, 5.2 points for the moderate CID, and 10.8 points for the large CID (Table 4).

UPDRS TOTAL SCORE

The study sample of 653 subjects with PD is described in Table 1. The sample was predominantly white, male, and married, with relatively high education and income.

UPDRS MOTOR SCORE

The mean UPDRS motor score (subscale III) was 27.2 (13.4). Based on the SE Scale ratings, the mean UPDRS motor scores ranged from a low of 15.3 (8.2) for the subjects reporting no disability (SE Scale, 100% [completely independent]) to a high of 60.0 (7.1) for subjects rated as totally dependent (SE Scale, 10% [bedridden]) (Table 2). Based on the HY stages, the mean UPDRS motor scores ranged from a low of 11.2 (4.9) for subjects with unilateral parkinsonism (stage 1) to a high of 54.4 (11.4) for subjects assessed as wheelchair bound or bedridden (stage 5) (Table 3). Regression analysis showed the average change on the UPDRS total score to be 8.6...
points for a 10% change on the SE Scale and 17.8 points for a 1-stage change on the HY stage. On the SF-12, 1 unit on the PH summary score was equivalent to a change of 0.85 points on the UPDRS total score, and 1 unit on the MH summary score was 0.91 points on the UPDRS total score. Therefore, a change of 1 SD (defined as 10 units) was 8.5 (PH) or 9.1 points (MH) on the UPDRS total score, and half of an SD was equivalent to 4.2 or 4.5 points, respectively (Table 4). The distribution-based analysis showed that the minimal CID was 4.1 points, the moderate CID was 10.7 points, and the large CID was 16.4 points on the UPDRS total score (Table 4). Based on a combination of the anchor- and distribution-based analyses, a change of 4.3 points is an appropriate estimate for the minimal CID, 9.1 points for the moderate CID, and 17.1 points for the large CID on the UPDRS total score (Table 4).

**Table 3. Cross-Sectional Analyses of Ratings on the UPDRS by HY Stages**

<table>
<thead>
<tr>
<th>HY Stagea</th>
<th>No. of Patients</th>
<th>UPDRS Motor Score</th>
<th>UPDRS Total Score</th>
<th>Δ From Previous Stageb</th>
<th>Δ From Previous Stageb</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean (SD) [Range]</td>
<td>Mean (SD) [Range]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-1.5</td>
<td>66</td>
<td>11.2 (4.9) [1-23]</td>
<td>16.6 (6.6) [7-36]</td>
<td>12.5</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>347</td>
<td>23.7 (9.3) [1-58]</td>
<td>35.5 (13.3) [4-87]</td>
<td>12.8</td>
<td>18.9</td>
</tr>
<tr>
<td>2.5</td>
<td>72</td>
<td>30.2 (10.5) [10-50]</td>
<td>45.3 (13.6) [17-75]</td>
<td>6.5</td>
<td>9.8</td>
</tr>
<tr>
<td>3</td>
<td>61</td>
<td>35.9 (9.3) [15-56]</td>
<td>54.7 (12.4) [33-86]</td>
<td>5.8</td>
<td>9.9</td>
</tr>
<tr>
<td>4</td>
<td>67</td>
<td>44.1 (12.7) [26-77]</td>
<td>69.1 (16.5) [29-106]</td>
<td>8.2</td>
<td>14.4</td>
</tr>
<tr>
<td>5</td>
<td>14</td>
<td>54.4 (11.4) [36-72]</td>
<td>90.2 (25.0) [64-118]</td>
<td>10.3</td>
<td>21.1</td>
</tr>
</tbody>
</table>

Abbreviations: Δ, change; HY, Hoehn and Yahr Scale (“on” for patients whose stage fluctuated); UPDRS, Unified Parkinson’s Disease Rating Scale.

a The regression coefficients were 10.8 (0.44) for the UPDRS motor score and 17.8 (0.61) for the UPDRS total score.
b All increments are significantly different at \( P < .001 \).

**Table 4. Summary of Distribution- and Anchor-Based Analyses of the CID on the UPDRS Total and Motor Scores**

<table>
<thead>
<tr>
<th>CID, UPDRS Motor Score</th>
<th>Minimal</th>
<th>Moderate</th>
<th>Large</th>
<th>Minimal</th>
<th>Moderate</th>
<th>Large</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distribution-based analysis</td>
<td>2.7</td>
<td>6.7</td>
<td>10.7</td>
<td>4.1</td>
<td>10.3</td>
<td>16.4</td>
</tr>
<tr>
<td>Anchor-based analysis</td>
<td>SF-12 PH</td>
<td>2.4</td>
<td>4.7</td>
<td>...</td>
<td>4.2</td>
<td>8.5</td>
</tr>
<tr>
<td></td>
<td>SF-12 MH</td>
<td>2.3</td>
<td>4.7</td>
<td>...</td>
<td>4.5</td>
<td>9.1</td>
</tr>
<tr>
<td></td>
<td>SE Scale</td>
<td>...</td>
<td>4.5</td>
<td>...</td>
<td>...</td>
<td>8.6</td>
</tr>
<tr>
<td></td>
<td>HY stages</td>
<td>...</td>
<td>...</td>
<td>10.8</td>
<td>...</td>
<td>17.8</td>
</tr>
<tr>
<td>Mean of all analyses</td>
<td>2.5</td>
<td>5.2</td>
<td>10.8</td>
<td>4.3</td>
<td>9.1</td>
<td>17.1</td>
</tr>
</tbody>
</table>

Abbreviations: CID, clinically important difference; ellipses, not applicable. For other abbreviations, see Table 1.

Concordance across a combination of distribution- and anchor-based approaches for analysis of the CID demonstrates that the moderate CID for the UPDRS motor score is approximately 5 points and for the UPDRS total score it is 9 points. Variability across sample populations and clinical settings suggests that a range of CID values is likely to be more useful than a single estimate.\(^{18,23}\) This study describes a range from minimal to moderate to large CID, corresponding to about 2.5, 5, and 11 points for the UPDRS motor score and 4.5, 9, and 17 points for the UPDRS total score.

The minimally clinically important change on the UPDRS was previously studied in a sample of individuals with early PD who participated in 2 clinical trials of dopamine agonist monotherapy.\(^9\) An anchor-based analysis using the Clinical Global Impression of Improvement found that the minimal clinically important change was 5 points on the UPDRS motor score and 8 points on the UPDRS total score.\(^9\) Our results show that the minimal CID was smaller: 2.5 points on the UPDRS motor score and 4.5 points on the UPDRS total score. Because the CID may vary at different stages of disease, this discrepancy may indicate that the CID is larger in earlier PD. Samsa et al\(^7\) questioned whether the initial decrement from perfect health to early symptoms may be more meaningful than the impact of similar decrements in the middle part of the scale. The discrepancies in the results between the present and earlier studies underscore the importance of replicating these analyses in different sample populations and in longitudinal studies. There were several differences between these 2 studies, including stage of disease, the anchor chosen, and the clinical setting (naturalistic vs clinical trial).

The presence of a larger CID in early PD is not supported by our analysis of UPDRS ratings by 10% decrements on the SE Scale (Table 2). If this were the case, one would anticipate larger increments in the UPDRS ratings between SE levels associated with earlier disability. However, the largest increment in the UPDRS motor score was associated with the change in SE Scale ratings from 30% to 20% in advanced PD, corresponding to the change...
in SE responses from “With effort, now and then does a few chores alone. . . . Much help needed” to “Nothing alone. Can be a slight help. . . . Severe invalid.” These larger UPDRS score increments associated with selected levels of 10% change on the SE Scale may simply identify SE Scale cut points that signal clinical distinctions more clearly. A single aberration is seen between SE Scale scores of 60% and 50%, where the UPDRS motor score goes down rather than up. The reason is unclear and requires further investigation.

Clinical trials in PD have applied arbitrary definitions of responders such as improvement of 20%, 30%, 3 points, and 5 points on the UPDRS motor scale. Three to 5 points on the UPDRS motor scale is precisely the minimal to moderate CID range in this analysis. Establishing the CID for a measure is particularly meaningful when this magnitude of improvement can be realistically achieved. In fact, the CIDs results in this study are consistent with effect sizes on the UPDRS found in recent clinical trials (Table 5). For example, in the Earlier vs Later Levodopa Therapy in Parkinson Disease study,26 in which 3 dosages of levodopa were studied (150, 300, and 600 mg/d), the change in UPDRS total score was 5.9 points for the low and moderate doses and 9.2 for the highest dose, corresponding to the minimal to moderate CID in this study. However, UPDRS score changes of 1 to 2 points for selegiline hydrochloride35 and rasagiline mesylate32 were below the minimal CID. Although the effect sizes in different trials described in Table 5 are not comparable owing to differences in study duration and variable adjustment for placebo, the range of effect sizes (1.1-8.4 for the UPDRS motor score and 1.8-9.2 for the UPDRS total score) are in the range of the CIDs found in this study. This study relied on a combination of patient-reported (SF-12) and clinician-reported (SE Scale and HY stage) assessments. Similarly, the UPDRS has elements of patient and clinician assessment, with subscales I (mentation, behavior, and mood) and II (activities of daily living) based on patient responses by history and subscale III (motor examination) based on clinical observation. The CID analysis was initially developed as a tool for patient-reported outcomes, particularly quality-of-life measures,7 and more recently has been applied to a greater diversity of measures, including physical performance measures.5,45

Clinically important differences may vary across diseases, ethnicity, and socioeconomic status. Therefore, these findings may not be applicable to patients who are nonwhite, have lower socioeconomic status, or have other forms of parkinsonism. The CID in a cross-sectional sample is conceptually distinct from CID analysis in a longitudinal sample, although cross-sectional and longitudinal analyses have yielded similar results in previous studies.7 The predetermined estimates of minimal, moderate, and large CID for each of 3 anchors may be subject to criticism as inaccurate representations of a clinically meaningful difference in PD. However, the results compare favorably with the calculated values of CID based on accepted standard effect sizes21 in the distribution-based analysis, and the ranges of computed values of CID based on each of the methods were in close agreement.

The minimum CID was originally defined as “the smallest difference in score in the domain of interest which patients perceive as beneficial and which would mandate, in the absence of troublesome side effects and excessive cost, a change in the patient’s management.”27 The role of a CID in clinical decision making is highlighted in this definition, although there is controversy about whether the CID is more applicable to the interpretation of group or individual differences. Indeed, assessing the risk to benefit ratio may be more straightforward on the individual level, where the treatment response may need to be especially robust to compensate for troublesome adverse effects or financial limitations. A range of CID values as demonstrated in this study (mini-
mal, moderate, and large) may help meet the needs of individual and group variability. For example, the use of the moderate to large CID range may be more suitable for interpreting change on the individual level, whereas the low end of the range (minimal to moderate) may be preferable when interpreting group differences.18

Establishing CID estimates for common outcome measures such as the UPDRS will not only influence patient management and clinical trials but also influence decision making by government and industry. Clinically important differences are a tool to aid clinicians in translating the results of statistically significant differences in large clinical trials to their individual patients. From a broader perspective, CIDs can facilitate the calculation of sample sizes in trials and may serve as a benchmark for interpreting treatment effects. This study shows that concordance of estimates of the CID on the UPDRS score can be achieved using multiple approaches of analysis and should be used to assess therapeutic interventions in PD.

Accepted for Publication: August 11, 2009.
Correspondence: Lisa M. Shulman, MD, Department of Neurology, University of Maryland School of Medicine, 110 S Paca St, Room 3-S-127, Baltimore, MD 21201 (lshulman@som.umaryland.edu).

Author Contributions: Study concept and design: Shulman, Gruber-Baldini, and Anderson. Acquisition of data: Shulman, Anderson, Fishman, Reich, and Weiner. Analysis and interpretation of data: Shulman, Gruber-Baldini, and Anderson. Drafting of the manuscript: Shulman, Gruber-Baldini, and Weiner. Critical revision of the manuscript for important intellectual content: Shulman, Gruber-Baldini, Anderson, Fishman, Reich, and Weiner. Statistical analysis: Gruber-Baldini. Obtained funding: Shulman and Weiner. Administrative, technical, and material support: Shulman, Fishman, and Weiner. Study supervision: Shulman, Anderson, and Weiner.

Financial Disclosure: None reported.

Funding/Support: This study was supported by the Rosalyn Newman Foundation.

Previous Presentation: This study was a platform presentation at the American Academy of Neurology Annual Meeting; May 1, 2007; Boston, Massachusetts.

REFERENCES


©2010 American Medical Association. All rights reserved.

45. Rascol O. Defining a minimal clinically relevant difference for the Unified Parkinson’s Rating Scale: an important but still unmet need. Mov Disord. 2006;21(8):1059-1061.