Movement Disorders
CLINICAL PRACTICE

Blinded Versus Unblinded Evaluations of Motor Scores in Patients with Parkinson's Disease Randomized to Deep Brain Stimulation or Best Medical Therapy

Gun-Marie Hariz, PhD,[1,] Anna Fredricks, RN,[1] Rasmus Stenmark-Persson, MD,[1,2] Marwan Hariz, MD, PhD,[1,3] Lars Forsgren, MD, PhD,[2] and Patric Blomstedt, MD, PhD[1]*

In deep brain stimulation (DBS) trials, blinded evaluation of patients using established disease-specific scales is considered gold standard for objective evaluation of results. Incidentally, an Editorial recently published in the BMJ stated that "re-evaluation of the role of blinding may be especially welcome in surgery, where it presents a particular challenge".[1] In an observer-blinded trial of DBS in the caudal Zona Incerta versus best medical management,[2] we aimed at evaluating possible differences between blinded and unblinded scores of the motor part of the Unified Parkinson's Disease Rating Scale (UPDRSIII).

Seventeen patients with Parkinson's disease (four women, age 58 ± 10 years, disease duration 9.6 ± 5.3 years) were scored after overnight medication withdrawal (off-med scoring) then again 1 hour after administration of a L-dopa dose 50% higher than the patient's normal morning dose (on-med scoring), and randomized to either DBS (eight patients) or best medical management (nine patients). At the 6 months follow up, the nine patients in the medical arm were scored again exactly as at baseline. The eight patients who had received DBS were scored in four conditions: off-medication on stimulation (off-med on-stim), off-medication off-stimulation (off-med off-stim), on-medication off-stimulation (on-med off-stim), and on-medication on-stimulation (on-med on-stim). The scorings were administered by an unblinded specialist nurse with 15 years experience of scoring UPDRS. The patients had a surgical cap covering the head during each scoring, and all scorings were videotaped.

The videotapes were presented in random order and without audio to two clinicians specialists in movement disorders (one neurologist and one neurosurgeon, with more than 30 years experience in the field) who were not involved in the selection of patients, or surgery, or follow up. When scoring UPDRSIII on videotapes, neither speech nor rigidity (items 18 and 22 of UPDRSIII) could be evaluated. Therefore, when comparing the blinded sum scores with unblinded ones, the values of these two items were excluded from the DBS nurse's sum scores. In case of discrepancies of scores between the two blinded clinicians, the videotape was viewed again and a consensus score was reached between them. Statistical analysis of the results was performed with SPSS. The Shapiro–Wilk test was used to test for normality. The Wilcoxon sign-rank test was used and results expressed in Median and InterQuartile range. Z-value was calculated. The level of significance was set to $P < 0.05$.

In total, 84 unblinded and 84 blinded UPDRSIII scorings were performed. Table 1 shows the scores according to the various medical and stimulation conditions: blinded vs. unblinded scores for patients undergoing L-dopa challenge showed no statistical difference, neither in off-med nor in on-med conditions. For patients who had DBS, there was a significant difference between scores only in off-med on-stim ($P = 0.026$) and in on-med on-stim ($P = 0.011$) conditions. Here, the median UPDRSIII scores rated by the unblinded nurse were 2.5 points lower than those rated by the blinded clinicians.

Our study showed a good agreement between the DBS nurse's unblinded scores and those of the blinded clinicians. The only significant differences concerned two conditions: off-med on- stim and on-med on-stim. Here the difference of the median values of UPDRSIII between evaluators amounted to only 2.5 points with almost similar interquartile range. Unsurprisingly, the

[1]*Unit of Deep Brain Stimulation, Department of Clinical Science, Neuroscience, Umeå University, Umeå, Sweden;* [2]*Department of Clinical Science, Neuroscience, Umeå University, Umeå, Sweden;* [3]*UCL Institute of Neurology, Queen Square, London, UK*

***Correspondence to:** Dr. Gun-Marie Hariz, PhD Department of Clinical Neuroscience, University Hospital, Umeå 901 85, Sweden; E-mail: gun.marie.hariz@umu.se

Keywords: Parkinson's disease, blinded trial, randomization, deep brain stimulation, UPDRS.
Relevant disclosures and conflicts of interest are listed at the end of this article.

Received 23 October 2020; revised 2 December 2020; accepted 6 December 2020.
Published online 13 January 2021 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/mdc3.13141

TABLE 1 UPDRS III scores in 17 patients in various clinical conditions

	Unblinded nurse Median (IQR)	Blinded clinician's Median (IQR)	Z	*P* value
A. 52 UPDRS III scores in 17 patients undergoing L-dopa challenge				
Off-medication	30.5 (12.5)	31.0 (15.50)	−.544	0.587
On-medication	12.0 (11.25)	12.5 (9.0)	−1.086	0.277
B. 32 UPDRS III scores in the 8 patients who received DBS				
Off-medication Off-stimulation	30.0 (20.25)	30.5 (18.25)	−.845	0.398
Off-medication On-stimulation	12.0 (12.0)	14.5 (13.5)	−2.226	0.026
On-medication Off-stimulation	12.0 (10.0)	12.5 (7.25)	−.511	0.610
On-medication On-stimulation	10.0 (4.75)	12.5 (6.0)	−2.539	0.011

Median and interquartile range (IQR) values are presented for: A. 52 scores in patients who underwent L-dopa challenge; B. 32 scores in eight patients with DBS.

unblinded DBS nurse, who met the patients on several occasions from baseline and onwards, tended to score more leniently when DBS was on. However, this 2.5 points difference falls below the cutoff score of 5 points deemed to represent a clinically meaningful difference.[3–5] Limitations of our study concern the low number of DBS patients, the video scorings of patients as opposed to live scoring, and the exclusion of speech and rigidity that could not be evaluated on video. Nevertheless, video scorings in blinded DBS trials are not uncommon,[2,5,6,7] and the use of standardized video for rating UPDRSIII is also well established in teaching the UPDRS.[8] Whether our study speaks or not in favor of obviating the need for blinding will require more research. What can be stated here is that provided an evaluator is well trained and has a long experience in scoring UPDRSIII in patients submitted to DBS, inter-rater variability between evaluators is low,[8,9] and the issue of blinding may not be vital to ascertain reliable and objective results.

Author Roles

(1) Research Project: A. Conception, B. Organization, C. Execution; (2) Statistical Analysis: A. Design, B. Execution, C. Review and Critique; (3) Manuscript Preparation: A. Writing of the first draft, B. Review and Critique.

G.M.H.: 1A, 1C, 2A, 2B, 3A
A.F.: 1B, 1C, 3B
R.S.P.: 2C, 3B
M.H.: 1A, 1C, 3B
L.F.: 1B, 2C, 3B
P.B.: 1A, 2C, 3B

Disclosures

Ethical Compliance Statement: We confirm that we have read the Journal's position on issues involved in ethical publication and affirm that this work is consistent with those guidelines. This study is a posthoc analysis of UPDRS scores from a randomized trial previously published in JNNP (please see reference 2), and approved by the local ethical committee of Umeå University. For the present work, informed patient consent was not necessary.

Funding Sources and Conflict of Interest: No specific funding was received for this work. The authors declare that there are no conflicts of interest relevant to this work.

Financial Disclosures for the Previous 12 Months: GMH, AF, RSP, and LF have nothing to disclose. MH: Honoraria and travel expenses from Boston Scientific for speaking at meetings. PB: Consultant for Abbott, Boston Scientific, Medtronic. Shareholder in Mithridaticum AB. ■

References

1. Godlee F. Blinding may be unnecessary but please divest. *BMJ* 2020;368: m255.
2. Blomstedt P, Stenmark-Persson R, Hariz GM, et al. Deep brain stimulation in the caudal zona incerta versus best medical treatment in patients with Parkinson's disease: a randomised blinded evaluation. *J Neurol Neurosurg Psychiatry* 2018;89:710–716.
3. Sánchez-Ferro Á, Matarazzo M, Martínez-Martín P, et al. Minimal clinically important difference for UPDRS-III in daily practice. *Mov Disord Clin Pract* 2018;5:448–450.
4. Schrag A, Sampaio C, Counsell N, Poewe W. Minimal clinically important change on the unified Parkinson's disease rating scale. *Mov Disord* 2006;21:1200–1207.
5. Weaver FM, Follett K, Stern M, et al. Bilateral deep brain stimulation vs best medical therapy for patients with advanced Parkinson disease: a randomized controlled trial. *JAMA* 2009;301:63–73.
6. Schuepbach WM, Rau J, Knudsen K, et al. Neurostimulation for Parkinson's disease with early motor complications. *N Engl J Med* 2013; 368:610–622.
7. Schjerling L, Hjermind LE, Jespersen B, et al. A randomized double-blind crossover trial comparing subthalamic and pallidal deep brain stimulation for dystonia. *J Neurosurg* 2013;119:1537–1545.
8. Goetz CG, Stebbins GT. Assuring interrater reliability for the UPDRS motor section: utility of the UPDRS teaching tape. *Mov Disord* 2004;19: 1453–1456.
9. Richards M, Marder K, Cote L, Mayeux R. Interrater reliability of the unified Parkinson's disease rating scale motor examination. *Mov Disord* 1994;9:89–91.