

Interrater Agreement in the Assessment of Motor Manifestations of Huntington's Disease

Penelope Hogarth,^{1*} Elise Kayson,² Karl Kieburtz,² Karen Marder,³ David Oakes,² Diana Rosas,⁴
Ira Shoulson,² Nancy S. Wexler,³ Anne B. Young,⁴ Hongwei Zhao,²
the United States–Venezuela Huntington's Disease Collaborative Research Group,
and the Huntington Study Group[†]

¹*Oregon Health and Science University, Portland, Oregon, USA*

²*University of Rochester, Rochester, New York, USA*

³*Columbia University, New York, New York, USA*

⁴*Massachusetts General Hospital, Boston, Massachusetts, USA*

Abstract: With prospects improving for experimental therapeutics aimed at postponing the onset of illness in preclinical carriers of the Huntington's disease (HD) gene, we assessed agreement among experienced clinicians with respect to the motor manifestations of HD, a relevant outcome measure for preventive trials in this population. Seventy-five clinicians experienced in the evaluation of patients with early HD and six non-clinicians were shown a videotape compiled from the film archives of the United States–Venezuela Collaborative HD Research Project. Observers were asked to rate a 2–3-minute segment of the motor examination for each of 17 at-risk subjects. The rating scale ranged from 0 (normal) to 4 (unequivocal extrapyramidal movement disorder characteristic of HD). As measured by a weighted κ statistic, there was substantial agreement among the 75

clinicians in the judgment of unequivocal motor abnormalities comparing scale ratings of 4 with ratings that were not 4 (weighted $\kappa = 0.67$; standard error (SE) = 0.09). Agreement among the non-clinicians was only fair (weighted $\kappa = 0.28$; SE = 0.10). Even under the artificial conditions of a videotape study, experienced clinicians show substantial agreement about the signs that constitute the motor manifestations of illness in subjects at risk for HD. We expect these findings to translate to a similar level of interobserver agreement in the clinical trial setting involving experienced investigators examining live patients. © 2004 Movement Disorder Society

Key words: Huntington's disease; clinical trials; disease progression

Historically, clinical trial efforts have been directed toward the evaluation of therapeutic interventions for manifest illness. In such trials, the efficacy of experimental therapeutics has been assessed typically using some measure of disease progression; for example, a change over time in Unified Parkinson's Disease Rating Scale

(UPDRS) score in Parkinson's disease. As increasing emphasis is placed on disease prevention by society and health care policy, a corresponding shift in the focus of clinical trials can be anticipated. With rational neuroprotective agents under development and the target at-risk population readily definable, Huntington's disease (HD) may represent an ideal model for this paradigm shift.

Although techniques employing volumetric magnetic resonance imaging (MRI) and positron emission tomography (PET) are under development as biomarkers to detect preclinical disease progression in HD, the clinical onset of illness is likely to remain a key endpoint for future clinical trials involving healthy at-risk subjects. Measuring the efficacy of a drug by its ability to delay illness onset will first require the systematic characterization and measurement of this clinical event that we

This article includes Supplementary Video Clips, available online at www.interscience.wiley.com/jpages/0885-3185/suppmat.

*Correspondence to: Dr. Penelope Hogarth, Oregon Health and Science University, 3181 SW Sam Jackson Park Road, OP32, Portland, OR 97239-3098. E-mail: hogarthp@ohsu.edu

[†]Members of the USVHDCRG and HSG are listed in the Appendix. Received 13 July 2000; Revised 2 May 2002; Accepted 14 February 2004

Published online 6 December 2004 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/mds.20332

have termed “phenoconversion.” Although the onset of HD could be defined in psychiatric or cognitive terms, these clinical features may lack specificity and could result in an unacceptably high false positive rate if used as clinical trial endpoints. The onset of a characteristic extrapyramidal movement disorder may offer a more specific and reliable endpoint for trials of putative neuroprotective agents in subjects at risk for HD. With such future trials in mind, we carried out experiments to assess interrater agreement regarding the motor manifestations of HD.

SUBJECTS AND METHODS

Seventeen subjects from the film archives of the US–Venezuela HD Collaborative Research Group¹ (USVH-DCRG) were selected and compiled on a videotape to illustrate a standardized examination of each subject. All subjects consented to a videotaped examination and were at immediate risk for HD by virtue of having a parent with the illness. The HD genetic status of all subjects remained concealed. Subjects were selected to encompass a full range of clinical normalities and abnormalities based on ratings given by the Venezuela project examiners and the videotape compilers. Subjects appeared on the videotape in random order.

The videotaped examinations included: (1) eye movements (saccades and smooth pursuits); (2) tongue protrusion; (3) finger taps; (4) pronation–supination of one hand on the other; (5) casual and tandem gait; and (6) the retropulsion pull test. There was also an opportunity to assess involuntary movements under conditions with subjects at rest and under stress (for example, performing a mental arithmetic task). A sample examination is included on the videotape that accompanies this article. The examination comprised most of the elements of the motor section of the Unified Huntington Disease Rating Scale (UHDRS), a comprehensive rating instrument developed by the Huntington Study Group (HSG) for use in clinical and research settings. The scale has been used in the evaluation of more than 4,000 at-risk and manifest HD subjects, and has been demonstrated to have excellent reliability and internal consistency.² A videotaped examination of tone was not felt to be interpretable and was edited from the tape. The Luria maneuver is not carried out as part of the examination in Venezuela and was thus excluded. The illustrated examination otherwise included all elements of the motor UHDRS.

Two separate experiments were conducted, involving a total of 75 clinician raters, all of whom were experienced in the assessment of patients with early HD, and 6 non-clinician raters who were knowledgeable about HD but had no clinical experience in the evaluation of pa-

tients or at-risk individuals. All raters were instructed to view the entire videotape at a single sitting, without replaying segments or discussing their ratings with colleagues. Observers were asked to score the 2–3-minute motor examination for each of the 17 subjects according to the motor UHDRS five-unit “diagnosis confidence level” shown in Table 1.

Attempts were made by the videotape compilers to include at least 3 subjects from each diagnosis confidence level in Table 1, plus an additional 2 subjects from diagnosis confidence level 4, the score that would represent an endpoint in a clinical trial setting. In the first experiment, all subjects rated as having an abnormal examination (any rating other than 0) by the videotape compilers were known to have later phenoconverted, or developed “unequivocal signs of HD.” The availability of sufficient good quality film archives prevented this same approach for the second experiment; however, all subjects rated as having unequivocal signs of HD (diagnosis confidence level 4) on the second videotape are known to have developed and retained a clinical diagnosis of HD.

In the first experiment, 35 clinician raters and 6 non-clinicians watched the videotape assembled together as a group, with no discussion or comparison of ratings allowed. In a separate experiment, a similar videotape compiled from the same archives was mailed to 40 different clinician raters who viewed the tape individually and assessed the motor examinations in the same manner.

Interrater agreement was assessed using κ coefficients, representing the proportion of agreement beyond that expected by chance.³ The prespecified analysis was the agreement between examiners for a rating of “4 versus not 4,” the designated threshold for the motor onset of

TABLE 1. Motor diagnosis confidence level taken from the Unified Huntington’s Disease Rating Scale

Rating	Definition
0	Normal (no abnormalities)
1	Nonspecific motor abnormalities (less than 50% confidence)
2	Motor abnormalities that may be signs of HD (50–89% confidence)
3	Motor abnormalities that are likely signs of HD (90–98% confidence)
4	Motor abnormalities that are unequivocal signs of HD ($\geq 99\%$ confidence)

Ratings to answer the question “To what degree are you confident that this person meets the operational definition of the unequivocal presence of an otherwise unexplained extrapyramidal movement disorder (e.g., chorea, dystonia, bradykinesia, rigidity) in a subject at risk for Huntington’s disease?” Guidelines for the use of the scale are provided to investigators through the Huntington Study Group.

HD in a therapeutic trial. In addition to an assessment of general agreement for all categories, interrater agreement was therefore assessed for a rating of "4 versus not 4" using a weighted κ , based on a formula by Schouten.⁴ The same weighted analysis was carried out to assess agreement for normal examinations ("0 versus not 0"). Because the videotape and viewing conditions were different for the two experiments, the results are analyzed and reported separately.

RESULTS

The results are summarized in Table 2. For the primary analysis of ratings "4 versus not 4" (HD versus not HD), interrater agreement was substantial between clinician raters, with a weighted κ of 0.67 and standard error (SE) of 0.09, coincidentally identical for both groups of clinician raters. Among the non-clinicians, agreement was fair for the same analysis, yielding a $\kappa = 0.28$ and $SE = 0.10$.

For ratings of "0 versus not 0" (normal versus not normal) agreement was moderate among clinician raters for both experiments (first group, $\kappa = 0.45$, $SE = 0.11$; second group; $\kappa = 0.42$, $SE = 0.07$). Non-clinicians achieved only fair agreement ($\kappa = 0.24$, $SE = 0.07$).

General agreement, taking into account all possible ratings (0, 1, 2, 3, and 4) and using nonweighted κ , was fair for both groups of clinician raters ($\kappa = 0.32$, $SE = 0.06$) and poor ($\kappa = 0.11$, $SE = 0.03$) for the non-clinicians.

DISCUSSION

This study demonstrates that trained and experienced clinicians show substantial agreement about the signs that constitute unequivocal motor manifestations of HD in subjects at risk for the illness, when using an examination tool with documented reliability and validity such as the UHDRS. Because the scale requires individual raters to be $\geq 99\%$ confident in making a diagnosis of

HD, one might have expected agreement to be even higher, particularly given the experience of the investigators involved in this study. That the κ achieved is not higher may be due related to the method of assessment, as watching a videotaped examination is a poor substitute for a clinical encounter with a live patient. The results also underscore the difficulty in assessing the earliest motor signs of HD. If motor onset of illness is to be used as an endpoint in future therapeutic trials in at-risk individuals, careful selection and training of the investigators performing the assessments will be necessary.

Several studies have investigated the clinical onset of HD. The US-Venezuela Collaborative HD Research Project, the largest prospective study involving individuals at risk for HD, was key to the identification of the mutant HD gene on the short arm of chromosome 4.⁵ In 1990, Penney and colleagues⁶ reported on the first 7 years of follow up on the individuals in the Venezuela kindred who were at nominal 50% risk for inheriting the HD gene. This report provided important information about the clinical abnormalities evident before phenocconversion, and identified those that best predicted illness onset within a defined period. Eye movement abnormalities and loss of dexterity in carrying out rapid alternating movements were the most consistent findings among at-risk subjects who later went on to develop manifest HD. The overall interrater agreement for definite motor onset of HD, using similar diagnostic criteria and involving similarly experienced investigators, showed a $\kappa = 0.61$, within the same range as our study.

DeBoo and colleagues⁷ found relatively poor agreement ($\kappa = 0.09-0.45$) between three neurologists who independently assessed the videotaped examination of 47 subjects of known gene status. Excellent agreement ($\kappa = 0.79, 0.90, \text{ and } 0.78$) was achieved only after the same neurologists discussed the cases, although false-positive diagnoses remained. The examination used in by deBoo and colleagues was not based on a validated scale, however, and the authors concluded that a reliable method of assessment would be crucial to future therapeutic trials.

It is worth pointing out that the current study is focused on reliability, not validity. The individual genetic status of the subjects on the videotape is unknown, so although raters agree about the signs that constitute a diagnosis, the accuracy of that diagnosis is not addressed. The study's applicability to future therapeutic trials in the at-risk population might therefore be questioned, because it is assumed widely that such trials would only be undertaken in subjects proven to carry the expanded form of the gene. However, if a putative neuroprotective agent were of low toxic potential (for ex-

TABLE 2. Interrater agreement among clinician raters and non-clinicians

	4 vs. not 4 (primary analysis)*	0 vs. not 0*	General agreement for all ratings
Experiment 1			
Clinicians (n = 35)	0.67 (0.09)	0.45 (0.11)	0.32 (0.06)
Non-clinicians (n = 6)	0.28 (0.10)	0.24 (0.07)	0.11 (0.03)
Experiment 2			
Clinicians (n = 40)	0.67 (0.09)	0.42 (0.07)	0.32 (0.06)

*Data reported as weighted κ ; values with standard error in parentheses except for general agreement values.

ample coenzyme Q10 or creatine), the ethical concerns about exposing gene-negative at-risk individuals would be mitigated, and a trial involving subjects blinded to their genetic status would become feasible. The fact that less than 5% of eligible individuals at risk for HD have chosen to undergo predictive testing⁸ suggests that there may be support among the at-risk community for a study design that allows a subject to remain unaware of their genetic status. The risk of harm from an unnecessary intervention must be weighed against the risk of harm from unwanted genetic knowledge. Certainly, precedent exists for interventional preventative clinical trials in which enrollment is based on estimated rather than known risk assessment for developing a disease. The Breast Cancer Prevention Trial (BCPT) enrolled 13,388 women calculated to have a 1.66% or greater 5-year risk for breast cancer and treated half of them with tamoxifen, a drug associated with a low but clearly increased risk of uterine cancer, stroke, and thromboembolic events.⁹ If this 1.66% 5-year risk of breast cancer was deemed worth the risks of treatment in the BCPT, it is not unreasonable to suppose that a 50% lifetime risk of HD would justify some risk of intervention in a future HD preventative trial.

Preventive clinical trials for HD may not be as remote as once thought. Multicenter clinical trials in subjects with manifest HD are underway or have been completed recently for the antiglutamatergic compounds riluzole¹⁰⁻¹² and remacemide,¹³ for coenzyme Q10¹³ and creatine,¹⁴ nutritional supplements with bioenergetic benefits, and for minocycline,¹⁵ an antibiotic with caspase-modifying actions. If these or other agents are ultimately found to slow the functional decline of subjects with manifest HD, their potential therapeutic application in preclinical at-risk subjects will become relevant. With prospects thus improving for neuroprotective interventions to delay the onset of HD, it is timely to define the methodology, feasibility, and ethical considerations of clinical research in this population. The HSG is currently addressing these issues in two NIH-supported multicenter studies involving 1,000 unaffected individuals at risk for HD in the Prospective Huntington At-Risk Observational Study (PHAROS; HG-02449) and 500 unaffected presymptomatic carriers of the HD gene in the Neurobiological Predictors of Huntington Disease (PREDICT-HD; NS-40068) study.

Acknowledgments: This work was supported by the Huntington's Disease Society of America, New York, NY, the Hereditary Disease Foundation, Santa Monica, CA, and United States Public Health Service (T32 NS 07338; Experimental Therapeutics in Neurological Disease), and the High Q Foun-

ation and the National Institutes of Health (HG 02449 and NS40098).

LEGEND TO THE VIDEO

A sample examination of a Venezuelan subject at risk for HD is shown.

APPENDIX

Huntington Study Group Participants

Charles Adler, Mayo Clinic, Scottsdale, AZ; Roger Albin, University of Michigan, Ann Arbor, MI; Tetsuo Ashizawa, Baylor College of Medicine, Houston, TX; Thomas Bird, University of Washington, Seattle, WA; Karen Blindauer, Medical College of Wisconsin, Milwaukee, WI; Scott Bundlie, Hennepin County Medical Center, Minneapolis, MN; James Caress, Wake Forest, Winston Salem, NC; John Caviness, Mayo Clinic, Scottsdale, AZ; Sylvain Chouinard, Montreal General Hospital, Montreal, Canada; Cynthia Comella, Rush-Presbyterian/St. Luke's Medical Center, Chicago, IL; Peter Como, University of Rochester, Rochester, NY; Jody Corey-Bloom, University of California-San Diego, San Diego, CA; Timothy Counihan, University of Rochester, Rochester, NY; Wallace Deckel, University of Connecticut, Hartford, CT; Richard Dubinsky, University of Kansas, Kansas City, KS; James Duffy, UCONN Huntington's Disease Program, Hartford, CT; Leon Dure, Children's Hospital of Alabama, Birmingham, AL; Stewart Factor, Albany Medical College, Albany, NY; Andrew Feigin, Manhasset, NY; North Shore University, Hubert Fernandez, Memorial Hospital of Rhode Island, Pawtucket, RI; Joseph Friedman, Memorial Hospital of Rhode Island, Pawtucket, RI; Timothy Greenamyre, Emory University, Atlanta, GA; Mark Guttman, Markham Health Centre, Toronto, Canada; Robert Hauser, University of South Florida, Tampa, FL; Steven Hersch, Emory University, Atlanta, GA; Bonnie Hersh, Boston University, Boston, MA; Susan Hickenbottom, University of Michigan, Ann Arbor, MI; Donald Higgins, Ohio State University, Columbus, OH; Douglas Hobson, Winnipeg Clinical, Winnipeg, Canada; George Jackson, UCLA Medical Center, Los Angeles, CA; Joseph Jankovic, Baylor College of Medicine, Houston, TX; Danna Jennings, Yale University, New Haven, CT; William Johnson, South Jersey Huntington's Disease Treatment Center, Stratford, NJ; William Koller, University of Miami, Miami, FL; Elan Louis, Columbia-Presbyterian Medical Center, New York, NY; Anne Louise Lafontaine, University of Calgary, Calgary, Canada; Carl Leventhal, Rockville, MD; William Mallonee, Hereditary Neurological Disease Center, Wichita, KS; Karen Marder, Columbia-Presbyterian Medical Center, New York, NY; Kenneth Marek, Yale University, New Haven, CT; Wayne Martin, University of Alberta, Edmonton, Canada; Marc Mentis, North Shore University Hospital, Manhasset, NY; Eric Molho, Albany Medical College, Albany, NY; Martha Nance, Hennepin County Medical Center, Minneapolis, MN; Christopher O'Brien, Colorado Neurological Institute, Englewood, CO; Constance Orme, University of Rochester, Rochester, NY; David Palmer, Hereditary Neurological Disease Center, Wichita, KS; Jane Paulsen, University of Iowa, Iowa City, IA; George Paulson, Ohio State University, Columbus, OH; Alan Percy, Children's Hospital of Alabama, Birmingham, AL; Susan Perlman, UCLA Medical Center, Los Angeles, CA; Joel Perlmutter, Washington University, St. Louis, MO; Gerald Podskalny, UMDNJ Robert Wood Johnson Medical Center, Stratford, NJ; Kimberly Quaid, Indiana University, Indianapolis, IN; Brad Racette, Washington University, St. Louis, MO; Juan Rachez-Ramos, University of South Florida, Tampa, FL; Lynn Raymond, University of British Columbia, Vancouver, Canada; David Richman, University of California-Davis, Sacramento, CA; Ted Roberts, University of Alberta, Edmonton, Canada; Robert Rofnitsky, University of Iowa, Iowa City, IA; Christopher Ross, Johns Hopkins University, Baltimore, MD; Guy Rouleau, Montreal General, Montreal, Canada; Allen Rubin, University of Pennsylvania, Philadelphia, PA; Daniel Sax, Boston University, Boston, MA; R. Neil

Schimke, University of Kansas, Kansas City, KS; Kathleen Shannon, Rush-Presbyterian/St. Luke's, Chicago, IL; Paul Shelton, Winnipeg Clinical, Winnipeg, Canada; Aileen Shinaman, University of Rochester, Rochester, NY; Leslie Shinobu, Massachusetts General Hospital, Boston, MA; Cliff Shults, University of California-San Diego, San Diego, CA; Eric Siemers, Indianapolis, IN; Oksana Suchowersky, University of Calgary Medical Center, Calgary, Canada; Francis Walker, Wake Forest University, Winston Salem, NC; William Weiner, University of Miami, Miami, FL; Vicki Wheelock, University of California-Davis, Sacramento, CA; Joanne Wojcieszek, Indiana University, Indianapolis, IN; Frederick Wooten, University of Virginia, Charlottesville, NC.

US-Venezuela Huntington's Disease Collaborative Research Group Participants

Ernesto Bonilla, University of Zulia, Maracaibo, Venezuela; Jang-Ho Cha, Massachusetts General Hospital, Boston, MA; Margot de Young, University of Zulia, Venezuela; Kenneth Fischbeck, National Institutes of Health, Bethesda, MD; Amerigo Negrette, University of Zulia, Venezuela; John B. Penney, Massachusetts General Hospital, Boston, MA; Maria Ramos, Hospital Virgen Del Camino, Spain; S. Robert Snodgrass, University of Mississippi Children's Hospital, Jackson, MS; Leslie Thompson, University of California, Irvine, Irvine, CA.

REFERENCES

1. Young AB, Shoulson I, Penney JB, et al. Huntington's disease in Venezuela: neurologic features and functional decline. *Neurology* 1986;36:244-249.
2. Huntington Study Group. Unified Huntington's Disease Rating Scale: reliability and consistency. *Mov Disord* 1996;11:136-142.
3. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas* 1960;20:37-46.
4. Schouten HJA. Measuring pairwise interobserver agreement when all subjects are judged by the same observers. *Stat Neerl* 1982;36:45-61.
5. Huntington's Disease Collaborative Research Group. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* 1993;72:971-983.
6. Penney JB, Young AB, Shoulson I, et al. Huntington's disease in Venezuela: 7 years of follow-up on symptomatic and asymptomatic individuals. *Mov Disord* 1990;5:93-99.
7. de Boo G, Tibben A, Hermans J, et al. Subtle involuntary movements are not reliable indicators of incipient Huntington's disease. *Mov Disord* 1998;13:96-99.
8. Nance MA, Myers RH, the US Huntington Disease Genetic Testing Group. Trends in predictive and prenatal testing for Huntington disease 1993-1999. *Am J Hum Genet* 1999;65:A406.
9. Fisher B, Costantino JP, Wickerham DL, et al. Tamoxifen for prevention of breast cancer: report of the National Surgical Adjuvant Breast and Bowel Project P-1 Study. *J Natl Cancer Inst* 1998;90:1371-1388.
10. Rosas HD, Koroshetz WJ, Jenkins BG, et al. Riluzole therapy in Huntington's disease. *Mov Disord* 1999;14:326-330.
11. Huntington Study Group. Dosage effects of Riluzole in Huntington's disease: a multicenter, placebo-controlled study. *Neurology* 2003;61:1551-1556.
12. Seppi K, Mueller J, Bodner T, et al. Riluzole in Huntington's disease (HD): an open label study with one year follow up. *J Neurol* 2001;248:866-869.
13. Huntington Study Group. A randomized, placebo-controlled trial of coenzyme Q10 and remacemide in Huntington's disease. *Neurology* 2001;57:397-404.
14. Huntington Study Group. A placebo-controlled trial of creatine in Huntington's disease. *Neurology* 2001;56:A386.
15. Huntington Study Group. Minocycline safety and tolerability in Huntington's disease. *Neurology* 2004; 63:547-549.