Assessment of the Performance of a Modified Motor Scale as Applied to Juvenile Onset Huntington's Disease

Mike C Horton[1], Peggy Nopoulos[2], Martha Nance[3], G Bernhard Landwehrmyer[4], Roger A Barker[5], Ferdinando Squitieri[6], REGISTRY Investigators of the European Huntington's Disease Network, Jean-Marc Burgunder[7], Oliver Quarrell[8].

1 Academic Department of Rehabilitation Medicine, Leeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds, Leeds, UK.

2 Departments of Psychiatry, Pediatrics, & Neurology, University of Iowa Carver College of Medicine, Iowa City, IA, USA.

3 Struthers Parkinson's Center, Minneapolis, MN, USA

4 Department of Neurology, University of Ulm, Ulm, Germany.

5 John Van Geest Centre for Brain Repair, Department of Clinical Neuroscience, Forvie Site, Robinson Way, Cambridge, CB2 0PY, UK.

6 Huntington and Rare Diseases Unit, IRCCS Casa Sollievo della Sofferenza Hospital (Rome-CSS Mendel), 71013 San Giovanni Rotondo, Italy.

7 Swiss HD Center, Neurozentrum Siloah and Department of Neurology, University of Bern, Switzerland

8 Department of Clinical Genetics, Sheffield Children's Hospital, Sheffield, UK.

Email:oliver.quarrell@sch.nhs.uk Telephone 441142717025 fax 441142737467

Running Title. Rasch Analysis JOHD Modified Motor Scale

REGISTRY Investigators of the European Huntington's Disease Network,

Registry Steering committee: Anne-Catherine Bachoud-Lévi, Anna-Rita Bentivoglio, Ida Biunno, Raphael M. Bonelli, Juliana Bronzova, Jean-Marc Burgunder, Stephen B. Dunnett, Joaquim J. Ferreira, Jan Frich, Joe Giuliano, Olivia J. Handley, Arvid Heiberg, Sergey Illarioshkin, Torsten Illmann, Jiri Klempir, G. Bernhard Landwehrmeyer, Jamie Levey, Tim McLean, Jørgen E. Nielsen, Susana Pro Koivisto, Markku Päivärinta, Sven Pålhagen, Oliver Quarrell, Maria Ramos-Arroyo, Raymund A.C. Roos, Carsten Saft, Ana Rojo Sebastián, Sarah J. Tabrizi, Wim Vandenberghe, Christine Verellen-Dumoulin, Tereza Uhrova, Jan Wahlström+, Jacek Zaremba

Language coordinators: Katrin Barth, Monica Bascuñana Garde, Kristina Becanovic, Sabrina Betz, Adrien Come, Leonor Correia Guedes, Jenny Callaghan, Selene Capodarca, Sébastien Charpentier, Martina Di Renzo, Daniel Ecker, Ana Maria Finisterra,, Ruth Fullam, Camille Genoves, Olivia J Handley, Andrea Horta, Carina Hvalstedt, Christine Held, Hasina Hussain, Kerstin Koppers, Claudia Lamanna, Matilde Laurà, Asunción Martínez Descals, Saul Martinez-Horta, Tiago Mestre, Sara Minster, Daniela Monza, Kristina Münkel, Lisanne Mütze, Martin Oehmen, Helene Padieu, Laurent Paterski, Nadia Peppa, Dawn Rogers, Ana Salgueiro, Catherine Taylor, Erika Timewell, Jenny Townhill, Patricia Trigo Cubillo, Grzegorz Witkowski, Abigail Wright, Daniel Zielonka, Eugeniusz Zielonka, Paola Zinzi

AUSTRALIA

Sydney (Department of Clinical Genetics, The Children's Hospital at Westmead): Fiona Richards

AUSTRIA

Innsbruck (Universitätsklinik Innsbruck, Neurologie): Eva-Maria Braunwarth, Florian Brugger, Lisa Buratti, Eva-Maria Hametner, Caroline Hepperger, Christiane Holas, Anna Hotter, Anna Hussl, Barbara Larcher, Philipp Mahlknecht, Christoph Müller, Bernadette Pinter, Werner Poewe, Eva-Magdalena Reiter, Klaus Seppi, Fabienne Sprenger, Gregor Wenning

BELGIUM

Charleroi (Institut de Pathologie et de Génétique (IPG)): Michel Dupuis, Cécile Minet, Pascale Ribaï, Dominique Van Paemel, Christine Verellen-Dumoulin

FRANCE

Angers (Centre de référence des maladies neurogénétique- CHU d'Angers): Philippe Allain, Dominique Bonneau, Marie Bost, Bénédicte Gohier, Marie-Anne Guérid, Audrey Olivier, Julie Prouzet, Adriana Prundean, Clarisse Scherer-Gagou, Christophe Verny

Paris (Hôpital de la Pitié Salpêtrière): Sandra Benaich, Alexis Brice, Sarah Boster, Perrine Charles, Alexandra Durr, Claire Ewenczyk, Hélène Francisque, Céline Jauffret, Damian Justo, Abdulrahman Kassar, Stephan Klebe, Fabien Lesne, Paolo Milani, Marie-Lorraine Monin, Tiffany Monnier, Emmanuel Roze, Alina Tataru, Maya Tchikviladzé

GERMANY

Bochum (Huntington-Zentrum (NRW) Bochum im St. Josef-Hospital): Jürgen Andrich, Gisa Ellrichmann, Lennard Herrmann, Rainer Hoffmann, Barbara Kaminski, Peter Kraus, Carsten Saft, Christiane Stamm

Hamburg (Universitätsklinikum Hamburg-Eppendorf, Klinik und Poliklinik für Neurologie): Kai Boelmans, Christos Ganos, Ines Goerendt, Walburgis Heinicke, Ute Hidding, Jan Lewerenz, Alexander Münchau, Michael Orth, Jenny Schmalfeld, Lars Stubbe, Simone Zittel

Münster (Universitätsklinikum Münster, Klinik und Poliklinik für Neurologie): Natalie Bechtel, Heike Beckmann, Stefan Bohlen, Nicole Göpfert, Eva Hölzner, Herwig Lange, Ralf Reilmann, Stefanie Rohm, Silke Rumpf, Christian Sass, Sigrun Schepers, Nathalia Weber

ITALY

Naples (Department of Neurosciences and Reproductive and Odontostomatological Sciences, Federico II University of Naples): Giuseppe De Michele, Luigi Di Maio, Carlo Rinaldi, Marco Massarelli, Silvio Peluso, Alessandro Roca, Cinzia Valeria Russo, Elena Salvatore, Pierpaolo Sorrentino, Tecla Tucci

Pozzilli (IS) (IRCCS Neuromed): Milena Cannella, Valentina Codella, Francesca De Gregorio, Annunziata De Nicola, Francesca Elifani, Chiara Esposito, Tiziana Martino, Irene Mazzante, Martina Petrollini, Maria Simonelli, Maurizio Vezza

IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo: Ferdinando Squitieri

Rome (LIRH Foundation): Barbara D'Alessio, Francesca Lovo

Rome (Azienda Ospedaliera Sant'Andrea; Department of Neuroscience, Mental Health and Sensory Organs (NESMOS), Faculty of Medicine and Psychology, Sapienza University of Rome; Institute of Translational Pharmacology & Institute of Cognitive Sciences and Technologies, National Research Council of Italy): Giulia Coarelli, Michela Ferraldeschi, Marina Frontali, Gioia Jacopini, Giovanni Ristori, Silvia Romano, Paola Zinzi

POLAND

Krakow (Krakowska Akademia Neurologii): Krzysztof Banaszkiewicz, Dorota Boćwińska, Kamila Bojakowska-Jaremek, Małgorzata Dec, Natalia Grabska, Malgorzata Krawczyk, Ewelina Kubowicz, Michalina Malec-Litwinowicz, Monika Rudzińska, Agata Stenwak, Andrzej Szczudlik, Elżbieta Szczygieł, Magdalena Wójcik, Anna Wasielewska

Poznan (Poznan University of Medical Sciences, Poland): Jacek Anioła Anna Bryl, Anna Ciesielska, Aneta Klimberg, Jerzy Marcinkowski, Husam Samara, Justyna Sempołowicz, Bartłomiej Wiśniewski, Daniel Zielonka

PORTUGAL

Lisbon-HSM (Hospital de Santa Maria, Clinical Pharmacology Unit, Instituto de Medicina Molecular): Leonor Correia Guedes, Miguel Coelho, Joaquim J Ferreira, Andreia Magalhães, Tiago Mestre, Tiago Mendes, Dulce Neutel, Filipe Rodrigues, Anabela Valadas

Porto-HGSA (Hospital Santo António- Centro Hospitalar do Porto): Gonçalo Cação, Sara Cavaco, Joana Damásio, Joana Fernandes, Alexandra Gonçalves, Rui Loureiro, Inês Moreira, Marina Magalhães, Paula Salgado

Porto- HSJ (Hospital de São João): Carlos Andrade, Andreia Costa, Carolina Garrett, Miguel Gago, Joana Guimarães, João Massano, Joana Meireles, Ana Monteiro

SPAIN

Burgos (Servicio de Neurología Hospital General Yagüe): Esther Cubo, Cecilia Gil Polo, Natividad Mariscal, Jesús Sánchez

Madrid RYC (Hospital Ramón y Cajal, Neurología): Mónica Bascuñana, Marta Fatás Ventura, Juan García Caldentey, Guillermo García Ribas, Justo García de Yébenes, José Luis López–Sendón Moreno, Verónica Mañanes Barral, Patricia Trigo Cubillo

SWEDEN

Umeå (Umeå University Hospital): Måns Berglund, Ghada Loutfi, Carina Olofsson, Eva-Lena Stattin, Laila Westman, Birgitta Wikström

SWITZERLAND

Bern: Jean-Marc Burgunder, Jessica Koehli, Yanik Stebler (Swiss HD Zentrum), Alain Kaelin, Irene Romero, Michael Schüpbach, Sabine Weber Zaugg (Zentrum für Bewegungsstörungen, Neurologische Klinik und Poliklinik, Universität Bern)

U.K.

Birmingham (The Barberry Centre, Dept of Psychiatry): Shahbana Akhtar, Jenny Crooks, Adrienne Curtis, Jenny de Souza (Keylock), John Piedad, Hugh Rickards, Jan Wright

Cambridge (Cambridge Centre for Brain Repair, Forvie Site): Roger A. Barker, Deidre O'Keefe, Anna Gerrtiz (nee Di Pietro), Kate Fisher, Anna Goodman, Susan Hill, Sarah Mason, Rachel Swain, Natalie Valle Guzman

Cardiff (Schools of Medicine and Biosciences, Cardiff University): Monica Busse, Cynthia Butcher, Jenny Callaghan, Stephen Dunnett, Catherine Clenaghan, Ruth Fullam, Sarah Hunt, Lesley Jones, Una Jones, Hanan Khalil, Sara Minster, Michael Owen, Kathleen Price, Jenny Townhill, Anne Rosser

Leeds (Chapel Allerton Hospital, Department of Clinical Genetics): Jeremy Cosgrove, Deena Gallantree, Stephanie Hamer, Emma Hobson, Stuart Jamieson , Alison Kraus, Mandy Longthorpe, Ivana Markova, Hannah Musgrave, Caroline Peacy, Ashok Raman, Liz Rowett, Jean Toscano, Sue Wild, Pam Yardumian

Leicester (Leicestershire Partnership Trust, Mill Lodge): Carole Clayton, Heather Dipple, Dawn Freire-Patino, Caroline Hallam, Julia Middleton

Manchester (Genetic Medicine, University of Manchester, Manchester Academic Health Sciences Centre and Central Manchester University Hospitals NHS Foundation Trust): Natalie Arran, Judith Bek, Jenny Callaghan, David Craufurd, Ruth Fullam, Marianne Hare, Liz Howard, Susan Huson, Liz Johnson, Mary Jones, Ashok Krishnamoorthy, Helen Murphy, Emma Oughton, Lucy Partington-Jones, Dawn Rogers, Andrea Sollom, Julie Snowden, Cheryl Stopford, Jennifer Thompson, Iris Trender-Gerhard, Nichola Verstraelen (formerly Ritchie), Leann Westmoreland

Sheffield (The Royal Hallamshire Hospital– Sheffield Children's Hospital): Oliver Bandmann, Alyson Bradbury, Helen Fairtlough, Kay Fillingham, Isabella Foustanos, Paul Gill, Mbombe Kazoka, Kirsty O'Donovan, Louise Nevitt, Nadia Peppa, Oliver Quarrell, Cat Taylor, Katherine Tidswell

ABSTRACT

Background: Huntington's disease can present at almost any age but traditionally, those with an onset ≤ 20 years are described as having juvenile onset Huntington's disease (JOHD). They are more likely to have bradykinesia and dystonia earlier in the course of the disease. The Total Motor Score of the Unified Huntington's Disease Rating Scale (UHDRS-TMS) is often used as the principal outcome measure in clinical trials.

Objective: To identify a motor scale more suitable for JOHD patients

Methods: A working group reviewed the UHDRS-TMS and modified it by adding four further assessment items. Rasch analysis was used to study the performance of the modified scale in 95 patients with a mean age of 19.4 (SD 6.6) years.

Results: The initial analysis showed a significant overall misfit to the Rasch model and a number of individual items displayed poor measurement properties: all items relating to chorea displayed significant misfit due to under-discrimination. Additionally, a number of items displayed disordered response category thresholds, and a large amount of dependency was present within the item set (96 out of 741 pairwise differences = 13%). An iterative process of scale re-structuring and evaluation was then undertaken, with a view to eliminating the largest sources of misfit and generating a set of items that would conform to Rasch model expectations.

Conclusion: This post-hoc scale restructuring appears to provide a valid motor score that is psychometrically robust in a JOHD population. This scale restructuring offers a pragmatic solution to measuring motor function in a JOHD population, and it could

provide the basis for the further iterative development of a more useful clinical rating scale for patients with JOHD.

INTRODUCTION

Huntington's disease (HD) is a progressive neurodegenerative disorder inherited as an autosomal dominant condition. It is characterised by a movement disorder, selective cognitive impairment and disturbance of affect. The disease results from an expansion of a CAG repeated sequence in the first exon of the *HTT* gene. A CAG repeat length of 36 -39 is associated with reduced penetrance whereas 40 or more is considered unequivocally abnormal. Onset may occur at almost any age but it is typically in midlife. [1][2][3][4] There is a negative correlation between age of onset and CAG repeat length, consequently those with JOHD often have a higher CAG repeat length.

The term juvenile onset HD (JOHD) is used to describe a patient having an age of onset of ≤ 20years: in well developed countries, this represents approximately 5% of cases [5]. This is not a distinct subcategory but remains a distinction because the phenotype is more likely to include bradykinesia and dystonia at an earlier stage of the illness, and little in the way of chorea. Clear problems arise with this definition because of the cut-off age of 20 years: firstly, it is arbitrary; a patient with onset at 18 years may not be significantly different from one with an age of onset of 22 years; secondly, a patient with onset under 10 years may have very different needs from a patient with onset as a late teenager; thirdly, a patient now aged 30 years could still have had age of onset ≤20 years.

Currently, there is no treatment to alter the natural history of the disease but treatment trials are under way for adult onset Huntington's disease [6] using standardised rating scales as outcome measures such as the Total Motor Score of the Unified Huntington's Disease Rating Scale (UHDRS) [7].

However, to ensure that the impact of any interventions being studied in clinical trials is assessed accurately, it is vital that the outcome measures adequately represent the constructs that the interventions are meant to treat or modify.

The UHDRS was developed and psychometrically validated using traditional approaches. However, modern psychometric assessment methodologies such as item response theory (IRT) [8] [9] [10] [11] and Rasch Measurement Theory (RMT) [12] offer various advantages over classical approaches, and can also reveal potential restrictions that may be present within these existing measures [13] [14].

Where traits cannot be measured directly, these are known as latent traits (e.g. depression, or quality of life). Latent traits are measured by indirect means, which is usually through multi-item scales. Rasch Measurement Theory provides a way to assess multi-item latent scales (i.e. Patient or Clinician-Reported Outcome Measures), to ensure that it is valid to add the individual items together to form an overall total score. The Rasch model is a unidimensional measurement model that satisfies the assumptions of fundamental measurement [15] [16], meaning that it provides a measurement template that scales can be tested against. The assumption that all items contribute independently to the total score is formally tested against the Rasch model, and any measurement anomalies within the item set are highlighted. The application of Rasch Measurement Theory provides a unified framework for several aspects of internal construct validity to be assessed. This includes the assessment of scale unidimensionality (whether all items are contributing to the same underlying construct), response category functioning (whether item response categories are working as they were intended to be used), response dependence (whether the response to any item has a direct implication to the response to any other item), scale targeting (relative distribution of item and

person locations), and item bias (whether an item is working in the same way for specific groups. e.g. males and females). When rating scale data conform to the Rasch model, the ordinal raw score can be transformed into a linear, interval scale, thus validating the use of parametric statistical procedures, although it should be noted that the raw score will remain ordinal.

The purpose of this study was to use a Rasch analytic framework to investigate the psychometric properties of a modified UHDRS motor scale on patients with JOHD. The UHDRS has already been assessed within an item response theory (IRT) framework [17], but the majority of subjects within that study had adult onset HD. Additionally, RMT offers a different approach to IRT. Whereas the IRT approach seeks to explain the variance in the data by adjusting the model to fit the data, in the RMT approach the model remains fixed, seeking to obtain invariant measurement by ensuring that items meet the requirements of the fixed model [18,19]

Furthermore, given the motor phenotype of JOHD, four items that quantified motor features common in JOHD were added to the UHDRS-TMS with a view to seeing if this helped the performance of the rating scale within a JOHD sample.

METHODS

A working group of the European Huntington's Disease Network met to consider the issue of assessing JOHD patients for intervention studies. Based on the opinion of an expert group, a strategic decision was made to retain the structure of the Total Motor Score scale, but an additional four items were added: 1) an overall chorea score: the current motor scale is heavily weighted towards chorea with an

assessment of seven areas (face, buccal, oral, lingual, trunk, right and left upper extremities) and as this clinical sign is less prominent in JOHD, an overall chorea may give a better assessment; 2) a repetitive hand-tapping task which quantifies bradykinesia; 3) a task timing how long it would take to drink 120 ml of water as another measure of capturing slowing of movements; and 4) a maximal tremor score because this has been suggested as a common clinical sign in JOHD. The modified scale has 39 items, each with five response categories, where a higher score represents a higher level of impairment; the additional items and responses are listed in Table 1.

This was a sub-study of the European Huntington's Disease Network REGISTRY project [20]: the ethical approval for REGISTRY included sub-studies. More information on the study can be found at http://www.euro-hd.net/html/registry [21]. Data from this extended JOHD motor assessment rating scale was collected from patients in Europe, the United States and Australia and submitted in paper form to the European Huntington's Disease Network REGISTRY site at Ulm and entered onto an Excel spread sheet.

Participants

Data was collected from 27 sites using multiple raters. Initially, there were 111 participants (58 females, 53 males) with a mean age of 21.3 years (SD 8.05, Range 6-40). Although it is possible for a person to develop HD as a late teenager and now be aged 40, the object of this study was to focus on the performance of the rating scale among younger patients; therefore, the analysis was restricted to those patients ≤30 years at the time of the study. This reduced the sample size to n=95.

Rasch Analysis of the data

The JOHD motor scale data was analysed using RUMM2030 software to investigate whether the pattern of item responses observed in the data matched the expectations of the Rasch measurement model [22]. The following fundamental aspects of the scale were assessed using this approach: overall fit to the model, adequacy of the response categories, individual item and person fit, local dependency, unidimensionality, differential item functioning (DIF), targeting of the scale and person separation reliability index (PSI). All of these elements have been previously described elsewhere [23] [24] [25]. Briefly, the data is compared to the fit assumptions of the Rasch model, so the tests-of-fit should be non-significant for the model assumptions to be satisfied. Individual items should demonstrate chi-square and ANOVA fit statistics >0.05 (Bonferroni adjusted), and the same ranges are applicable for any DIF tests. A residual correlation (Q3) value of 0.2 was used to indicate dependency. This is slightly more lenient than the value of 0.2 above the average correlation that is currently recommended [25], but it was felt that this was a reasonable compromise, given the low sample size involved. A series of t-tests were used to assess unidimensionality [26], where evidence of unidimensionality is apparent when the lower bound 95% CI percentage of significantly different t-tests is < 5%.

RESULTS

The 95 patients in this study had a male:female ratio of 45:50. The mean age was 19.4 years (SD 6.6 years). The age distribution was 6 – 30 years: 13 were aged 6-10 years; 34 were aged 11-20 years and 48 were ages 21- 30 years. The mean CAG repeat length was 67.6 (SD 15.6). The median CAG repeat length was 63 (inter-quartile range 56-77) with a range of 46-117.

Analysis

The initial analysis showed a significant overall misfit to the Rasch model (see Table 2, Analysis 1), and a number of individual items displayed poor measurement properties. At this point, the key observation was that all items relating to chorea (both the original 7-area assessment and the overall chorea score) displayed significant misfit due to under-discrimination. This suggests that the chorea items are not usefully contributing to the total score of the item set due to a lack of discrimination within this patient group. An example of this under-discrimination is presented in Figure 1. Additionally, a number of items displayed disordered response category thresholds (see Figure 2), and a large amount of dependency was present within the item set (96 out of 741 pairwise differences = 13%). It was noted that the majority of the dependency was clustering into groups of items that were related to the same concept; for example, dystonia in the right and left upper limb.

An iterative process of scale re-structuring and evaluation was then undertaken, with a view to eliminating the largest sources of misfit and generating a set of items that would conform to Rasch model expectations. The summary results of each analysis stage are presented in Table 2.

Analysis 2 Restructuring of the data to reduce dependency.

The first stage of data reconstruction was carried out in order to reduce the impact of item dependency. This was done by creating a number of composite items, which considered the maximum level of impairment shown within a dependent cluster of items. This approach conserves the informative clinical information, whilst taking account of the psychometric impact of inter-item dependency. Practically, this meant that the 'right' and 'left' elements of the items: pronate/supinate hands, rigidity of arms, finger taps, bradykinesia, hand-tapping, tremor upper extremity and tremor lower extremity were combined into single items, with the highest value of the right and left elements selected for the composite item. Similarly, the assessment of maximal dystonia was reduced from five elements down to three, with the retention of 'trunk' as a single item, but combining the 'right' and 'left' elements relating to the upper and lower extremity. Additionally, the assessments of eye movements were restructured to combine the 'horizontal' and 'vertical' elements of the ocular pursuit, saccade initiation, and saccade velocity items.

None of the chorea items appeared to discriminate across the level of motor function in the initial analysis, and all elements clustered in terms of the evidence of dependency. To investigate whether the under-discrimination remained present with a single maximal chorea value, all of the original 7 assessments of chorea were combined into a single maximal chorea item, selecting the highest observed value across all elements. The global chorea item was also retained individually at this stage.

These amendments resulted in the reduction of the item set from 39 to 22 items, each with five available response categories. These amendments are represented

as 'restructure run 1' in Table 3, and the fit statistics are presented as Analysis 2 in Table 2.

Analysis 3: Further restructuring of the data to remove dependency.

A degree of response dependency was still present at this point, leading to further restructuring. This included the combination of all elements of saccade initiation and velocity into a single composite item, all elements of dystonia into a single composite item, and all elements of tremor into a single composite item. These further amendments resulted in the reduction of the item set to 18 items, each with five available response categories. These amendments are represented as 'restructure run 2' in Table 3, and the fit statistics are presented as Analysis 3 in Table 2.

Analysis 4: Removal of chorea due to under-discrimination

At this stage, both the newly created chorea composite item and the original global chorea item continued to display a large degree of misfit due to under-discrimination. Both of these items were duly removed from the item set, therefore reducing the scale to 16 items. The fit statistics at this stage are presented as Analysis 4 in Table 2.

Analysis 5: Collapsing of response categories due to disordered response.

After accounting for the majority of the response dependencies and removing the clear anomalies of the chorea items, six items were still displaying disordered

response categories, indicating that they did not seem to be functioning in the intended manner. With guidance from the patterns of disorder and expert input from an experienced HD clinician, the response categories of these six items were collapsed. In the case of the items relating to the retropulsion pull test, maximal dystonia and Luria tri-step score, the response categories were reduced from five to four. In the case of the items relating to tongue protrusion, maximal rigidity of arms and bradykinesia of drinking, the response categories were reduced from five to three. Although this is presented as a single analysis stage, all rescoring was carried out iteratively. The rescoring structure is presented in Table 4, with the pre and post-rescoring response category threshold plots presented in Figure 3. The fit statistics at this stage are presented as Analysis 5 in Table 2.

Analysis 6 & 7 – removal of tremor and dysarthria items

Following the rescore, the 'max tremor' composite item was still displaying a misfit due to an under-discrimination (similar to the chorea items). Additionally, the 'dysarthria' item was also displaying a misfit, although this item was indicating an over-discriminating response pattern in addition to displaying a lower-level dependency across a number of items, which seemed to be adversely influencing the dimensionality of the item set. The 'max tremor' and 'dysarthria' items were iteratively removed from the item set, and the fit statistics are presented as Analysis 6 and 7 respectively, in Table 2.

Following the removal of these two items, the overall fit was good (chi-square p=0.297), individual item fit was good, dimensionality was acceptable (7/95 = 7.37% [95% CI = 3-11.8%]), all response categories were ordered, and there was no DIF by

gender, handedness, or the version of the scale that was administered (adult or JOHD-specific). The relative person-item distribution (targeting) plot also indicated that the remaining item set still captured the full range of impairment, meaning that the scale reliability statistics remained high (see Figure 4). However, some dependency remained apparent between the two items measuring balance (retropulsion pull and tandem walk) (0.316), and a lower-level dependency between the 'max rigidity arms' and 'max dystonia items' (0.238).

Discussion

To our knowledge, this is the first attempt to assess the motor scale of the UHDRS on a population of people with JOHD. The motor scale revealed areas of significant dependency between clusters of items relating to different elements of common concepts, including a consistent dependency between the right and left side of impairments to motor function. A *post-hoc* correction was applied to account for the apparent dependency, which also went some way to alleviating the apparent lack of unidimensionality within the item set. At a practical level, this reduces the amount of items that are summed to form the total score, but the key clinical information relating to the location of impairment is retained. Given the nature of the observed dependencies within the item set, it is reasonable to expect a similar finding to be observed within the UHDRS motor scale when applied to patients with adult onset HD.

Additionally, it appears as though the five-category response options are not working as intended across all items. This can mean that respondents are not differentiating between certain responses, and therefore the response choice between these categories becomes arbitrary. This may be the case where too many response options are presented, or if response category labelling is unclear, inconsistent, or difficult to quantify. For example, this appeared to be the case with the 'tongue protrusion' item, where respondents appeared to have difficulty distinguishing between response categories labelled such as 'cannot fully protrude tongue' and 'cannot protrude tongue beyond lips'. In this case, the analysis has highlighted an issue which is made clearer when put into context. As these two response categories are not mutually exclusive, it is perhaps not surprising that respondents are not clearly differentiating between the available response options. Again, a post-hoc

adjustment was made to the scale by collapsing some of the response categories of the items displaying disordered thresholds. Although this improved scale fit, it should be noted that the actual response options (as presented) have not changed. The process of collapsing merely treats certain response options as equivalent; the practical application of this is described below.

Furthermore, it was recognised from the outset of the original scale development, that the motor scale was heavily weighted towards chorea, and for this reason the global chorea score was introduced. Within the reference frame of the JOHD sample, it was shown that none of the items relating to chorea were usefully discriminating across different levels of the trait, including the new global chorea score. It is to be expected that a worsening chorea score would be associated with a worsening overall motor score, but this was not found to be the case in the JOHD sample. This has significant implications for clinical trials because the UHDRS-TMS is heavily weighted towards chorea. The objective of current research is to provide disease modifying treatments. The recent work by Schobel et al to modify the principal outcome measure is helpful but still uses theUHDRS-TMS in the formula [27] in chorea.

Although the amended (collapsed) bradykinesia drinking item appeared to work well within the final item set, the feedback from scale users has raised concerns regarding this item. These concerns relate to the practical aspect associated with drinking 120ml of water, as this may create a potential choking hazard. This item should therefore be considered carefully within any future scale development.

Of the remaining new items: the assessment of tremor did not prove useful because it was removed at the analysis stage; however the bradykinesia hand tapping assessment was retained.

The major limitation of this study was the sample size that was available for analysis (n=95). A sub-analysis to investigate the effect of CAG repeat length on the performance of the scale could not be undertaken. However, given the relative rarity of JOHD, it took an enormous amount of time and resource to collect a sample of this apparently modest size, so this sample size should be considered in context. It is acknowledged that this small sample size may result in more instability surrounding the analysis results. However, the magnitude of the results and the replication of the key findings suggest a larger sample would deliver largely equivalent results. A further limitation is that age of onset was not collected systematically as data were collected separately from the main REGISTRY data, so we were not able to stratify the results into those with onset ≤10 years from those with onset in their late teens. We tried to mitigate this limitation by restricting the study to those ≤30 years of age.

Conclusion

This study was designed to assess the use of a modified motor scale in JOHD. The modifications from the adult-oriented UHDRS motor scale were modest, as it was thought that making additional assessments would result in the scale being more focused for the JOHD population. This study has revealed some significant implications for assessing future treatments in young people with HD. The analysis suggests that it is not valid to sum the 39 items (35 original, and 4 newly added) of the UHDRS-JOHD motor scale into a single total score.

On the basis of this result the HD research community could consider the following options: (1) Continue using the UHDRS-TMS and accept its limitations; (2) Retain the UHDRS-TMS, but additionally include the new hand tapping task; (3) Retain the UHDRS-TMS plus the hand tapping task so that it does not look any different to the practising clinician undertaking an intervention study. However, it will be necessary to pre-specify that any outcomes will be based on post-hoc changes to the scale, as described in this study; (4) Redesign the JOHD motor rating scale from the ground up, using qualitative and quantitative methods to iteratively develop a motor rating scale which is clinically meaningful.

Ideally a rating scale should focus on functional assessments rather than an amalgam of clinical signs. Although option 4 would perhaps present the best solution, in the short term we would recommend option 3 because it has the practical merit of being easily adopted without significant further work. However, option 4 should be considered as a future direction in which to take this informative work, which is presented as part of an iterative development process.

Conflicts of Interest

The authors have no conflicts of interest to report in relation to this work

REFERENCES

[1]. Walker FO. Huntington's disease. Lancet (London, England). 2007 Jan;369(9557):218–28.

[2]. Novak MJU, Tabrizi SJ. Huntington's disease. BMJ. 2010 Jun;340:c3109.

[3]. Bates GP, Dorsey R, Gusella JF, Hayden MR, Kay C, Leavitt BR, et al. Huntington disease. Vol. 1, Nature reviews. Disease primers. England; 2015. p. 15005.

[4]. Newcombe RG. A life table for onset of Huntington's chorea. Ann Hum Genet. 1981 Oct;45(4):375–85.

[5]. Quarrell O, O'Donovan KL, Bandmann O, Strong M. The Prevalence of Juvenile Huntington's Disease: A Review of the Literature and Meta-Analysis. PLoS Curr. 2012 Jul;

[6]. Rodrigues FB, Wild EJ. Clinical Trials Corner: September 2017. J Huntingtons Dis. 2017;6(3):255–63.

[7]. Kieburtz K. Unified Huntington's Disease Rating Scale: reliability and consistency. Huntington Study Group. Mov Disord. 1996 Mar;11(2):136–42.

[8]. Lord F. A theory of test scores. Psychom Monogr [Internet]. 1952 [cited 2017 Dec 22];7(7). Available from: https://www.psychometricsociety.org/sites/default/files/pdf/MN07.pdf

[9]. Lord FM. The relation of the reliability of multiple-choice tests to the distribution of item difficulties. Psychometrika [Internet]. 1952 Jun [cited 2017 Dec 22];17(2):181–94. Available from: http://link.springer.com/10.1007/BF02288781

[10]. Lord, Frederic M. ,Novick MR. Statistical Theories of Mental Test Scores. Oxford, England: Addison-Wesley; 1968.

[11]. Hambleton RK, Swaminathan H. Item response theory : principles and applications. Boston, MA: Kluwer-Nijhoff; 1985.

[12]. Rasch G (Georg). Probabilistic models for some intelligence and attainment tests. University of Chicago Press; 1980.

[13]. Hobart J, Cano S. Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. Health technology assessment (Winchester, England). England; 2009.

[14]. Yorke J, Horton M, Jones PW. A critique of Rasch analysis using the Dyspnoea-12 as an illustrative example. J Adv Nurs. 2012 Jan;68(1):191–8.

[15]. Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? Arthritis Rheum. 2007 Dec;57(8):1358–62.

[16]. Newby VA, Conner GR, Grant CP, Bunderson CV. The Rasch model and additive conjoint measurement. J Appl Meas. 2009;10(4):348–54.

[17]. Vaccarino AL, Anderson K, Borowsky B, Duff K, Giuliano J, Guttman M, et al. An item response analysis of the motor and behavioral subscales of the unified Huntington's disease rating scale in huntington disease gene expansion carriers. Mov Disord [Internet]. 2011 Apr [cited 2017 Dec 22];26(5):877–84. Available from: http://doi.wiley.com/10.1002/mds.23574

[18] Andrich D. Rating scales and Rasch measurement. Expert review of pharmacoeconomics & outcomes research. 2011 Oct;11(5):571-85. PubMed PMID: 21958102.

[19] Andrich D. Controversy and the Rasch model: a characteristic of incompatible paradigms? Med Care. 2004 Jan;42(1 Suppl):I7-16. PubMed PMID: 14707751.

[20]. Orth M, Handley OJ, Schwenke C, Dunnett SB, Craufurd D, Ho AK, et al. Observing Huntington's Disease: the European Huntington's Disease Network's REGISTRY. PLoS Curr. 2010 Sep;

[21]. European HD Registry [Internet]. [cited 2018 Apr 10]. Available from: http://www.euro-hd.net/html/registry

[22]. Andrich D. Rasch models for measurement. Sage Publications; 1988.

[23]. Hagquist C, Bruce M, Gustavsson JP. Using the Rasch model in nursing research: an introduction and illustrative example. Int J Nurs Stud. 2009 Mar;46(3):380–93.

[24]. McNally G, Rickards H, Horton M, Craufurd D. Exploring the Validity of the Short Version of the Problem Behaviours Assessment (PBA-s) for Huntington's disease: A Rasch Analysis. J Huntingtons Dis. 2015;4(4):347–69.

[25]. Christensen KB, Makransky G, Horton M. Critical Values for Yen's Q_3: Identification of Local Dependence in the Rasch Model Using Residual Correlations. Appl Psychol Meas [Internet]. 2017 May 16 [cited 2017 Dec 22];41(3):178–94. Available from: http://journals.sagepub.com/doi/10.1177/0146621616677520

[26]. Smith EVJ. Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. J Appl Meas. 2002;3(2):205–31.

[27]. Schobel SA, Palermo PA, Long JD, et al. Motor, cognitive, and functional declines contribute toa single progessive factor in early HD. Neurology 2017; 89(24):2495-2502.

Table 1. Summary of the four additional items used in this study

Chorea Global

    0 = absent

    1 = slight/intermittent

    2 = mild/common or moderate/intermittent

    3 = moderate/common

    4 = marked/prolonged

Bradykinesia – hand tapping

    0 = > 100 in 30 seconds

    1 = 80-100 in 30 seconds

    2 = 60-79 in 30 seconds

    3 = 40-59 in 30 seconds

    4 = < 40 in 30 second

Bradykinesia – drinking

    0= <5 seconds to drink 120ml of water

    1 = 5-7 seconds to drink 120ml of water

    2 = >7-11 seconds to drink 120ml of water

    3 = >11 to 18 seconds to drink 120ml of water

    4 = >18 seconds to drink 120ml of water

Total time to drink 120 ml of water

Maximal tremor

    0 = absent

    1 = slight/intermittent

    2 = mild/common

    3 = moderate/common

    4 = marked and severe

Table 2 – Analysis stage summary fit details

Analysis Number	Description	no. of items	n	Item Fit Residual		Person Fit Residual		Chi Square Interaction			PSI	Alpha	Unidimensionality % significant t-tests (lower bound 95%CI)	no. of pairwise dependencies >0.2
				Mean	SD	Mean	SD	Value	df	P				
1	Initial analysis	39	95	-0.18	1.51	-0.2	1.55	261.7	78	>0.0005	0.95	0.95	35.8% (31.4%)	96/741 (13%)
2	dependency reduction run 1	22	95	0.03	1.84	-0.06	1.19	184.8	44	>0.0005	0.94	0.94	21.05% (16.7%)	19/231 (8%)
3	dependency reduction run 2	18	95	0.02	2.04	-0.06	1.19	186.1	36	>0.0005	0.93	0.93	18.95% (14.6%)	11/153 (7%)
4	remove both chorea items	16	95	0.24	1.74	0.06	1.05	86.84	32	>0.0005	0.94	0.94	10.53% (6.1%)	6/120 (5%)
5	rescore items	16	95	0.04	1.58	-0.01	1.02	64.24	32	0.0006	0.94	0.94	10.53% (6.1%)	5/120 (4%)
6	remove max tremor item	15	95	0.06	1.27	-0.06	0.9	38.59	30	0.135	0.94	0.94	9.47% (5.1%)	2/105 (2%)
7	remove dysarthria item	14	95	0.08	0.99	-0.06	0.84	31.45	28	0.297	0.94	0.94	7.37% (3.0%)	2/91 (2%)

Table 3. Restructure framework

Item number	Descriptor	Item number under restructure run:		New item
		1	2	
1	ocular pursuit horizontal	1	1	max ocular pursuit
2	ocular pursuit vertical			
3	saccade initiation horizontal	2	2	max saccade
4	saccade initiation vertical			
5	saccade velocity horizontal	3		
6	saccade velocity vertical			
7	dysarthria	4	3	dysarthria
8	tongue protrusion	5	4	tongue protrusion
9	finger taps right	6	5	max finger taps
10	finger taps left			
11	pronate/supinate hands right	7	6	max pronate/supinate hands
12	pronate/supinate hands left			
13	Luria tri step test	8	7	Luria
14	rigidity-arms right	9	8	max rigidity arms
15	rigidity-arms left			
16	bradykinesia - body	10	9	bradykinesia - body
17	maximal dystonia trunk	11	10	max dystonia
18	maximal dystonia RUE	12		
19	maximal dystonia LUE			
20	maximal dystonia RLE	13		
21	maximal dystonia LLE			
22	maximal chorea face	14	11	max chorea
23	maximal chorea BOL			
24	maximal chorea trunk			
25	maximal chorea RUE			
26	maximal chorea LUE			
27	maximal chorea RLE			
28	maximal chorea LLE			
29	gait	15	12	gait
30	tandem walking	16	13	tandem walking
31	retropulsion pull test	17	14	retropulsion pull test
32	chorea – global	18	15	chorea - global
33	bradykinesia – hand tapping right	19	16	max bradykinesia – hand tapping
34	bradykinesia – hand tapping left			
35	bradykinesia - drinking	20	17	bradykinesia - drinking
36	maximal tremor RUE	21	18	max tremor
37	maximal tremor LUE			
38	maximal tremor RLE	22		
39	maximal tremor LLE			

Table 4. Rescoring structure

Item	Statement	Max Score	Rescored	Original response category				
				1	2	3	4	5
1	max ocular pursuit	4		0	1	2	3	4
2	max saccade	4		0	1	2	3	4
3	dysarthria	4		0	1	2	3	4
4	tongue protrusion	2	Y	0	1	1	2	2
5	max finger taps	4		0	1	2	3	4
6	max pronate/supinate hands	4		0	1	2	3	4
7	Luria tri step test	3	Y	0	1	1	2	3
8	max rigidity arms	2	Y	0	1	1	2	2
9	bradykinesia - body	4		0	1	2	3	4
10	max dystonia	3	Y	0	1	1	2	3
12	gait	4		0	1	2	3	4
13	tandem walking	4		0	1	2	3	4
14	retropulsion pull test	3	Y	0	1	2	3	3
16	max bradykinesia – hand tapping	4		0	1	2	3	4
17	bradykinesia - drinking	2	Y	0	1	1	2	2
18	max tremor	4		0	1	2	3	4

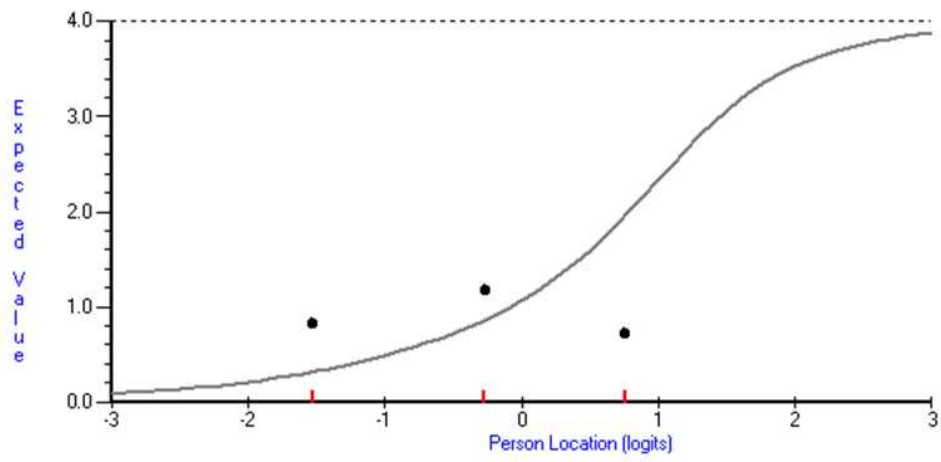**Figure 1. Item Characteristic Curve of the under-discriminating Chorea item**

Figure 2. Example of items with ordered thresholds (Max. Finger Taps –
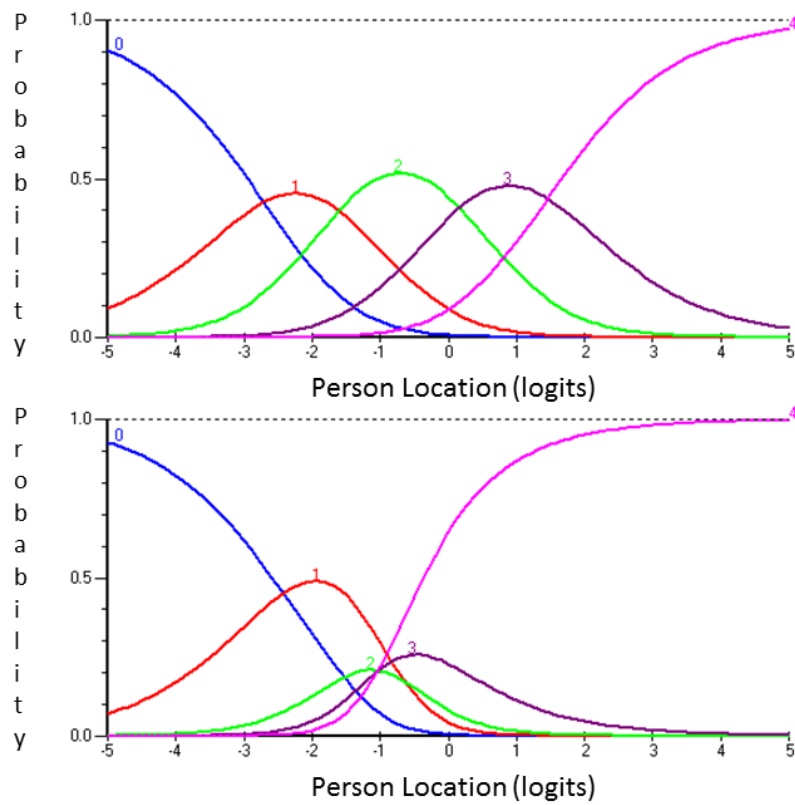**upper plot) and disordered thresholds (Bradykinesia Drinking – lower**
plot)

Figure 3. Item threshold map pre and post-rescoring

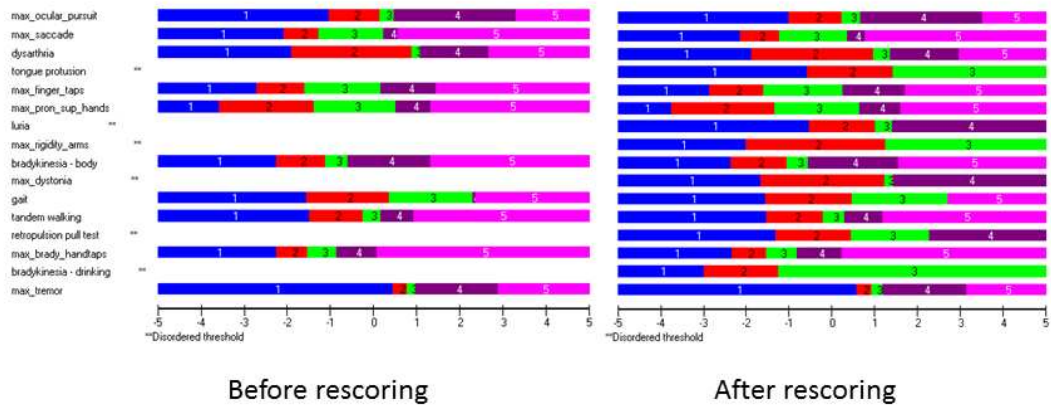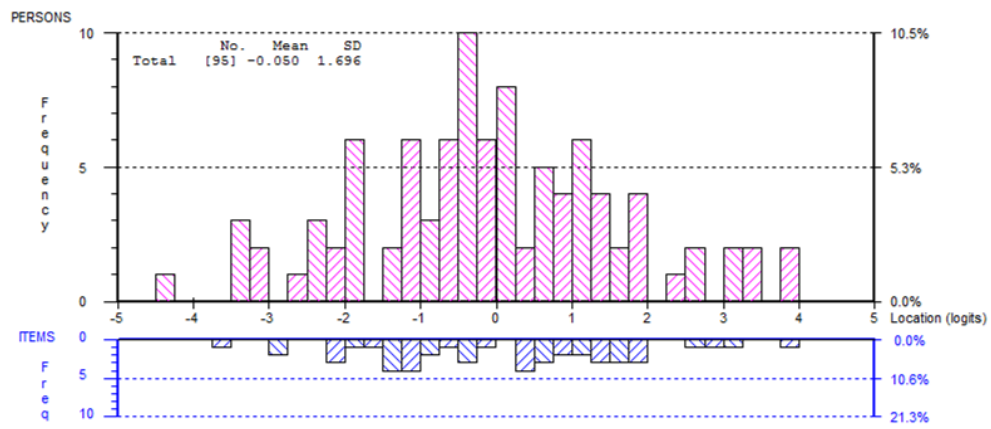Before rescoring   After rescoring

Figure 4. Relative person-item threshold distribution of final 14 items

Legends to Figures

Figure 1

The grey line represents the expected response curve, and the black dots represent the observed data. The 'flat' nature of the curve suggests that respondents are obtaining an approximately equal value on the item, regardless of their underlying level of motor function (as represented by the x-axis), and is therefore not discriminating across the level of motor function.

Figure 2

Each curve represents the inferred probability distribution of persons responding in a particular response category, given their underlying level of motor function. Each response category should emerge at some point as the most likely response on the underlying scale. In turn, all response category thresholds should therefore be ordered (see upper plot). Where response categories do not function as intended, response category thresholds become disordered (see lower plot).

Figure 3

Plot shows the relative location of all response category thresholds across all items. Where no plot is shown for an item, the thresholds were disordered.

Figure 4

Plot shows the relative distribution of logit locations of all item thresholds (below x-axis) and persons (above x-axis). These distributions should align where a scale is well-targeted to the population being measured.