UNIVERSITY OF LEEDS

This is a repository copy of *Exploring the Validity of the Short Version of the Problem Behaviours Assessment (PBA-s) for Huntington's disease: A Rasch Analysis.*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/96137/

Version: Accepted Version

Article:
McNally, G, Rickards, H, Horton, M et al. (1 more author) (2015) Exploring the Validity of the Short Version of the Problem Behaviours Assessment (PBA-s) for Huntington's disease: A Rasch Analysis. Journal of Huntington's disease, 4 (4). pp. 347-369. ISSN 1879-6397

https://doi.org/10.3233/jhd-150164

Reuse
Unless indicated otherwise, fulltext items are protected by copyright with all rights reserved. The copyright exception in section 29 of the Copyright, Designs and Patents Act 1988 allows the making of a single copy solely for the purpose of non-commercial research or private study within the limits of fair dealing. The publisher or other rights-holder may allow further reproduction and re-use of this version - refer to the White Rose Research Online record for this item. Where records identify the publisher as the copyright holder, users can verify any specific terms of use on the publisher's website.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

White Rose university consortium
Universities of Leeds, Sheffield & York

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Exploring the Validity of the Short Version of the Problem Behaviours Assessment (PBA-s) for Huntington's disease: A Rasch Analysis

1st author: George McNally[a]

2nd author: Hugh Rickards[b]

3rd author: Mike Horton[c]

Last author: David Craufurd[d]

[a]College of Medical and Dental Sciences, University of Birmingham, UK.

[b]Neuropsychiatry Department, The Barberry, Birmingham, UK.

[c]Academic Department of Rehabilitation Medicine, Leeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds, Leeds, UK.

[d]Faculty of Medicine and Human Sciences, Institute of Human Development, University of Manchester and Manchester Academic Health Science Centre, Manchester, UK.

Running Title: Rasch analysis of the PBA-s

Word Count

Abstract: 240 words

Manuscript: 5017 words

Manuscript Reference Number: 15-164 (1st revised submission)

All authors have contributed the work, agree with the presented findings, the work has not been published before nor is being considered for publication in another journal.

Correspondence address

Brooke House,
16 Church St,
Braunston,
Rutland,
UK,
LE15 8QT

Email Address: gem060@bham.ac.uk
Telephone number: 07742089855

Abstract

Background: The short version of the Problem Behaviours Assessment (PBA-s) is the recommended outcome measure for behavioural symptoms in Huntington's disease. Rasch analysis was used to further investigate the measurement limitations of the PBA-s.

Objectives: 1) To assess the psychometric properties of the 11 severity and frequency items within the PBA-s and 2) to determine the construct validity of using a total PBA-s score as a clinical outcome measure.

Methods: PBA-s data for 517 participants from Enroll-HD were included in the Rasch analysis. Separate analyses were conducted for the severity and frequency items of the PBA-s, using RUMM2030 software. Achieving fit to the model provides supporting evidence that all items contribute to a single underlying latent trait. This property is defined as internal construct validity.

Results: The total PBA-s severity score demonstrated several important limitations, including disordered response categories for all 11 severity items, local dependency and poor targeting. However, modifying the original five-point scoring system to a four-point system resulted in ordered response categories for seven of the severity items and achieved a good overall fit to the Rasch model. For the total PBA-s frequency score, fit to the model was not achieved even after amendments to the scoring system.

Conclusions: This study suggests that with reduction to a four-point scoring system, the total PBA-s severity score may be considered a valid clinical outcome measure. This study also suggests limitations in the use of a total PBA-s frequency score.

Keywords: Rasch, Psychometrics, Huntington's disease, Problem Behaviours Assessment, Enroll-HD

Introduction

Huntington's disease (HD) is an autosomal dominant neurodegenerative disorder characterised by the development of debilitating motor, cognitive and behavioural symptoms [1]. The genetic defect responsible for HD is an unstable and extended CAG repeat length on the gene that encodes huntingtin, located on chromosome 4 [2]. In the Western world, the prevalence of HD is 4-12 per 100,000 and onset usually occurs insidiously in the fourth or fifth decade of life [3,4]. Behavioural (neuropsychiatric) symptoms in HD have been reported to have a greater impact on quality of life [5] and functional disability [6] than either cognitive or motor symptoms. These manifestations may develop many years before the onset of distinctive motor signs and are increasingly recognised as the main reason for institutionalisation, due to the disabling and distressing impact they have on both patients and their carers [6-8], making their early recognition vital.

A wide range of behavioural symptoms have been recognised in HD, most commonly depression, irritability, apathy and anxiety which occur in up to seventy-six percent of patients [9]. Obsessive-compulsive behaviours, suicidal ideation and psychosis (hallucinations and delusions) occur less frequently. The short version of the Problem Behaviours Assessment (PBA-s) is a semi-structured interview containing 11 items, each designed to measure the severity and frequency of a different behavioural symptom in HD [1]. The PBA-s was developed by the EHDN Behavioural Working Group from an original 40-item version (PBA-HD), however, the shorter interview is more commonly used in clinical practice, as well as in multicentre RCTs and international observational studies such as REGISTRY and ENROLL-HD. In a recent systematic review it was identified as the recommended rating scale for behavioural symptoms in HD [10].

Traditionally, the method used for the development and evaluation of rating scales has been classical test theory (CTT) [11]. This approach focuses mainly on person-level statistics, such as means and standard deviations. CTT also uses test-level statistics such as Cronbach's

alpha to assess reliability and factor analysis to determine the dimensionality of a measurement [12,13]. Factor analyses performed on the short and long version of the PBA have reported consistent findings, providing evidence to support its reliability from a traditional psychometric perspective [1,14,15].

However, modern psychometric techniques such as Rasch analysis are increasingly adopted as a means to further investigate limitations in the use and interpretation of clinical outcome measures [16]. The Rasch approach provides an in-depth understanding of a rating scale's measurement properties and allows for identification of measurement issues not detected by CTT analyses [17]. The Rasch measurement model applies the assumption that scores obtained for individual symptoms vary with respect to the overall severity of the trait being measured. For example, a person with more advanced behavioural symptoms would be more likely to score highly on items indicating a more severe clinical picture, such as suicidal ideation. The pattern of item responses in the sample data are tested against the expectations of the Rasch model. Achieving fit to the model provides supporting evidence that all items contribute to a single underlying latent trait, such as the overall severity of behavioural symptoms. This property is defined as internal construct validity [17,18].

Rating scales are increasingly used as the primary outcome measure in clinical trials for neurological diseases, making them the main dependent variables that influence decisions made about the efficacy of future treatments and patient care [19]. With recent advances towards disease-modifying agents for HD [20], Rasch analysis provides a platform for reducing the risk of type-1 and type-2 errors in trials due to poor quality rating scales. Improving the ability of the PBA-s to detect important symptom changes has the potential to influence the outcome of trials investigating the effectiveness of new treatments for the disabling and debilitating behavioural symptoms of HD. The Rasch model has been used to successfully evaluate other psychiatric rating scales, including the Hamilton Rating Scale for Depression [13], the Hospital Anxiety and Depression Scale [18] and the Beck Depression Inventory [21].

The aim of this study was to determine the validity of the PBA-s for Huntington's disease using Rasch analysis. More specifically, the analysis process aimed to 1) assess the psychometric properties of the 11 severity and frequency items within the PBA-s and 2) determine the construct validity of using a total PBA-s score as a clinical outcome measure. To our knowledge, this is the first study to perform Rasch analysis on the PBA-s.

Materials and Methods

Study design

This study analysed retrospective anonymised data from Enroll-HD, an ongoing multicentre longitudinal observational study established primarily in Europe and North America to monitor disease progression. Access to the Enroll-HD database allowed for generalisation to an international population of HD patients, maximising the external validity and minimising the confounding effects of cultural and national differences in the presentation of behavioural symptoms assessed by the PBA-s [22].

Participants

Since Enroll-HD was established in July 2012, participants have been continuously recruited into the database. Participants were required to visit their respective site to gather data on motor, cognitive and behavioural symptoms at baseline. Eligible participants were asked to attend annual follow-up visits. Individuals with choreic movement disorders that were negative for the HD mutation were not eligible for inclusion in Enroll-HD.

For this study, demographic data for participants recruited between July 2012 and January 2015 were obtained from Enroll-HD on age, sex and CAG repeat length. Participants with a positive HD genotype were eligible for inclusion in this study, defined by the presence of ≥ 36 CAG repeats [23]. Subjects with a normal CAG repeat length of ≤ 27 or intermediate

repeat range of 27-35 were excluded from the study [24]. Participants with a positive HD genotype who had not yet met the motor-defined diagnostic criteria (premanifest HD), as judged by a qualified interviewer, were included alongside participants with motor signs (manifest HD). The premanifest HD group were included as neuropsychiatric symptoms can occur several years before the onset of motor signs [7,8]. Total Functional Capacity (TFC) score (0-13) was used to classify manifest HD participants into five categories, lower scores signifying greater functional impairment [25]. Additionally, participants with a CAG repeat length ≥ 55 were excluded from this study, as this is strongly correlated with juvenile Huntington's disease, which presents with a different phenotype [26]. No randomisation process was required for this study and all subjects that met the eligibility criteria were included in the sample population.

Access to the Enroll-HD data was approved by the Enroll-HD Scientific Publication Review Committee. All sites were approved by local ethics committees in their respective countries, ensuring written informed consent was signed by willing participants that fully understood the requirements of inclusion in Enroll-HD. This study was granted ethical approval from the local Internal Ethics Review Committee, University of Birmingham.

Assessment of neuropsychiatric symptoms

The PBA-s is an 11-item semi-structured interview specifically designed to address the most common behavioural and psychiatric symptoms of HD [1]. Each item is structured to elicit information from the patient, in order to enable the interviewer to categorise the patient into one of five pre-defined rating categories relating to each behavioural symptom. The short version of the PBA is recommended in clinical practice and performed more commonly in comparison to the original 40-item interview [10].

The five-point PBA-s rating scale (0-4) includes two subscales for severity and frequency (Table 1), modelled on the previously recommended behavioural section of the

Unified Huntington's Disease Rating Scale (UHDRS) [27]. The PBA-s assesses symptoms over the past four weeks, improving the recall of events by patients and the accuracy of the interview in comparison to the behavioural section of the UHDRS, which attempted to rate behaviour over the previous six months [1]. Only assessments completed in English were included in this study to avoid any inconsistencies of how items perform in different languages.

The PBA-s is conducted either in clinic or at home, ideally in the presence of a knowledgeable informant, for example a relative or paid carer. The informant and patient are given the opportunity to speak to the interviewer together and separately. Discussion with the informant may elicit additional insight into the subject's behaviours, which could not easily be obtained in their presence [28]. The trained interviewer is then required to make a clinical judgment with regard to each item score, taking all information and observations into account. For this study, only scores completed by qualified interviewers, in the presence of the subject and an informant, were used to achieve a high quality data set. The PBA-s has been shown to have substantial agreement in scores given by different interviewers, indicating a good inter-rater reliability [15].

Data analysis

Descriptive statistics were used to display the clinical characteristics and demographics of the study population.

The PBA-s data were analysed using RUMM2030 software [29] to investigate whether the pattern of item responses observed in the data matched the expectations of the Rasch measurement model [17]. All 11 severity item scores were summed together to create a total PBA-s severity score and the same method was used to produce a total PBA-s frequency score. Rasch analysis enabled the construct validity of these two total PBA-s scores to be evaluated. Separate analyses were carried out for the severity scale and the frequency scale, but the analytic procedure was replicated for each scale.

Due to the consistent polytomous structure (i.e. more than two response categories) of the PBA-s, the initial step in Rasch analysis was to conduct a likelihood ratio test. This determined which mathematical derivation of the Rasch model was more appropriate for the data set. A significant result for the likelihood ratio test ($p<0.05$) supports the use of the partial credit model [30] instead of the simpler rating scale model [31].

The following fundamental aspects of Rasch analysis were assessed:

1) *Overall fit to the model*: this was evaluated using the total chi-square item-trait interaction statistics for both the severity and frequency dimensions of the PBA-s [18,32]. A non-significant chi-square probability value gives an indication of a good level of overall fit, using a Bonferroni alpha value adjusted for the number of items [33]. The item-person interaction statistics summarise the individual item fit and person fit to the model. These standardised fit residual values approximate a z-score, and therefore a perfect fit would result in a mean value of 0 and a standard deviation of 1 [18]. These summary residual statistics (and the deviation from the perfect values) may give an overall impression of the fit, although these do not reveal specific item-level and person-level misfit.

2) *Adequacy of the response categories*: threshold maps and category probability curves were examined to identify disordered thresholds as a potential cause of misfit [32]. A threshold is the point between two adjacent response categories when the probability of the respondent endorsing either option is 50% (e.g. equally likely to score a “1” or a “2”) [12]. A disordered threshold indicates that a response category is never the most likely response, at any underlying level of the trait in question. This implies that the original response categories are not functioning as intended, and this may be due to a number of reasons [18], including that assessors find it difficult to differentiate between the various response categories for that particular item. When disordered response categories were encountered, categories were collapsed together and rescored to correct for the apparent disorder and improve fit to the model.

3) *Individual item and person fit:* standardised fit-residual values for items and persons were examined for any indication of misfit (values outside of ± 2.5). The residual value is the deviation from the Rasch model, summated for each individual item or person [18]. Individual item chi-square fit statistics were also assessed, using a Bonferroni-adjusted alpha-level.

4) *Local dependency:* defined by the response to any one item being dependent on the response to any other item, after controlling for the underlying trait. To investigate local dependency between items, a residual correlation value of more than 0.2 above the average of all item residual correlations was considered indicative [34].

5) *Unidimensionality:* to determine whether the scale was measuring a single unidimensional construct, principal component analysis (PCA) of the residuals was conducted to identify the two most different subsets of items (i.e. the most positively and negatively factor loading items on the first component). T-tests were performed comparing the scores on the two subsets of items for each person in the sample [35]. If more than 5% of the t-tests were significant (more specifically if the lower 95% confidence limit exceeded 5%), the scale was not considered unidimensional.

6) *Differential item functioning (DIF)*: a form of item bias that can occur when different groups within the sample (e.g. males and females), despite equal levels of the underlying trait, respond differently to an item. DIF was examined for each item with respect to age (dichotomised at a median of 55 years), gender and time-point (baseline or follow-up) using analysis of variance with a Bonferroni-adjusted alpha-level [18]. When one subgroup (e.g. females) consistently score differently on an item, across all levels of the trait, this is known as uniform DIF; when DIF varies across levels of the trait, this is known as non-uniform DIF.

7) *Targeting of the scale*: assessed by comparing the mean location score for persons with the mean value of zero set for the difficulty of the items. For a well targeted scale, the mean location for persons would be close to zero, indicated by inspection of the person-item threshold distribution map [18,32].

8) *Person separation reliability index (PSI):* examined to assess the internal consistency reliability of the scale and the ability of the measure to discriminate amongst persons with different levels of the underlying trait. Interpretation is comparable to Cronbach's alpha coefficient where minimum values of 0.7 and 0.85 indicate acceptable reliability for group and individual use, respectively [32].

Some participants in the database had completed both baseline and follow-up PBA-s interviews. To eliminate the possibility of time-series dependency (as some characteristics of the participants crossed both time-points), a sample was created so that each participant was only included in the data set once and each time-point was represented equally [36]. In this study, a total of 517 participants were included, which was optimal for Rasch analysis as larger samples inflate the chi-square fit statistics, which may falsely suggest misfit [37].

Results

Sample

A total of 822 participants obtained from Enroll-HD were eligible for inclusion in this study (Figure 1). The mean age of the sample was 54.7 years (range, 20.6-87.7 years) and 48.4% were males. Five participants were excluded due to incomplete data and 300 participants were excluded due to the lack of a knowledgeable informant at the baseline and annual follow-up interview. Of the remaining 517 participants, 266 participants were only interviewed with an informant at baseline and were included in the sample. 251 different participants were interviewed with an informant at baseline and follow-up. To eliminate time-series dependency, only follow-up data were included for these participants so each time-point was equally represented. The final sample consisted of 89 (17.2%) participants with premanifest HD and 428 (82.8%) participants with manifest HD, within which the Total Functioning Capacity stage ranged from I-V. The clinical characteristics and demographics of the sample are given in Table 2.

Rasch analysis

The likelihood ratio test was significant ($p<0.001$) for both the frequency and severity dimensions of PBA-s, supporting use of the partial credit model in this study.

PBA-s severity items

Initial analysis of the 11-item PBA-s severity score revealed a non-significant chi-square item-trait interaction statistic ($\chi^2=111.3$, df=77, $p=0.0065$), indicating borderline fit to the model after using a Bonferroni-adjusted p-value (Table 3, Analysis 1). Summary fit residual SDs for items (SD=1.28) and persons (SD=0.50) were within acceptable limits. All individual item and person fit residuals were within ± 2.5 and individual item chi-square probabilities were non-significant.

Inspection of the category probability curves demonstrated disordered response thresholds for all 11 items. The curves indicated that assessors had difficulty differentiating between response category 1 “slight, questionable” and 2 “mild (present, not a problem)” on the original five-point scale (Figure 2). However, creating a four-point scoring system for all items by collapsing these two response categories into a single category resulted in ordered thresholds for seven of the items, improving fit to the model ($\chi^2=84.8$, df=66, $p=0.059$; Table 3, Analysis 2). This is graphically illustrated by figure 2, showing that as the level of trait increases, each response category in turn has a point along the level of trait when it is the most likely response category to be endorsed.

Four items still displayed disordered thresholds after creating a four-point scoring system. However, item 5 “angry or aggressive behaviour” only displayed marginal disorder, therefore the item was not deleted. For item 2 “suicidal ideation”, item 9 “delusions” and item 10 “hallucinations”, the frequency of responses in categories above 0 “absent” were low, with percentages of 9.40%, 5.73% and 1.83% respectively. Deletion of the three items did not improve model fit (Table 3, Analysis 3) and given their clinical relevance, the three items were also retained.

Local dependency was detected between item 1 “depressed mood” and item 2 “suicidal ideation” with a residual correlation of r=0.14, more than 0.2 above the average residual correlation of r=-0.086. Item 4 “irritability” and item 5 “angry or aggressive behaviour” also displayed local dependency (r=0.17). Grouping each pair of dependent items in sub-test analysis accounted for the dependency between the items. The PSI decreased from 0.55 to 0.50, but overall interpretation of the fit statistics was the same as previously. To maintain the integrity of the scale, none of the dependent items were removed and no further action was taken. Additionally, dependency can influence response thresholds, therefore the disordered thresholds for item 2 and item 5 may be in part due to their local dependency.

Testing for dimensionality revealed that the proportion of significant t-tests was 5.05%, outside the critical value of 5.00%. However, the 95% CI lower bound was 3.00%, supporting the concept that the total PBA-s severity score measures a single unidimensional construct.

No item bias (DIF) with respect to age or time-point was detected. However, significant uniform DIF for gender on item 1 “depressed mood” was detected, after applying a Bonferroni-adjusted alpha-level (Figure 3). Inspection of the item characteristic curve revealed that item 1 was biased towards females, indicating that despite equal levels of underlying trait, females were more likely to endorse the item than males. Grouping of item 1 with two other items (item 3 “anxiety” and item 6 “apathy”) displaying marginal uniform DIF (detected at a 5% alpha level) in subtest analysis, revealed that the DIF for gender cancelled out at the overall scale level [38]. Hence, no further action was taken.

Inspection of the person-item distribution map (Figure 4) revealed that the scale was not well targeted (mean persons location was -2.00; SD=0.873). The easiest items to endorse were item 4 “irritability” and item 6 “apathy”, whereas the most difficult items to endorse were item 9 “delusions” and item 10 “hallucinations”. The PSI with and without extreme values were 0.55 and 0.51 respectively, and the Cronbach’s Alpha (α) value was 0.75, however, the Alpha value is likely to be artificially high as it does not take targeting into account. These low values are likely to have been affected by the poorly targeted skewed distribution.

PBA-s frequency items

The chi-square item-trait interaction statistic for the 11-item PBA-s frequency score was significant (χ^2=156.26, df=77, p<0.001), indicating poor fit to the model (Table 3, Analysis 4). Summary and individual fit residuals and SDs for items and persons were within acceptable limits. However, all items displayed disordered thresholds and individual item chi-square probabilities were significant for three items (1, 5 and 11), indicating item misfit.

Despite attempts at collapsing to four-point and three-point scoring systems for all items (based on the semantics of the original response category descriptions), the thresholds remained largely disordered. Ordered thresholds were only achieved for items 1,3-5, and 7 after creating a three-point scoring system (Table 3, Analysis 5) by collapsing “seldom” and “sometimes” (scores of 1 and 2) and “frequently” and “daily” (scores of 3 and 4).

Similar to the severity items, local dependency existed between item 1 “depressed mood” and 2 “suicidal ideation”, as well as for item 4 “irritability” and 5 “angry or aggressive behaviour”.

No DIF for age, sex or time-point was detected and the scale showed no evidence of multidimensionality. After collapsing to a three-point scale, the chi-square item-trait interaction was still significant (p<0.001), the PSI with and without extremes were low (0.54 and 0.49 respectively) and the mean persons location was -1.38 (SD = 0.83), indicating poor targeting.

Discussion

To our knowledge, this is the first study to investigate the validity of the PBA-s as a clinical outcome measure using Rasch analysis. The PBA-s is the recommended outcome measure for behavioural symptoms in HD, as it has been recognised as a reliable and valid tool using traditional psychometric techniques [10,15]. However, applying Rasch analysis has revealed

a number of limitations when using total PBA-s scores. This was demonstrated by disordered thresholds, local dependency, and poor targeting of the scale in this study.

The findings of this study have shown that modifying the structure of the PBA-s improved its construct validity. The total PBA-s severity score demonstrated good fit to the Rasch model after modification to a four-point scale. However, fit to the model was not achieved for the total PBA-s frequency score even after reduction to a three-point scale.

Originally, the PBA-s was created to assess neuropsychiatric symptoms individually and was not designed to produce a total score. However some studies have multiplied the severity and frequency scores for each item and subsequently added these individual item scores to create subscale scores [14,39]. For example, symptom clusters derived from previous factor analyses have led to the creation of composite affect, irritability and apathy subscales [1,14,15]. Broader multiplicative behaviour scores have been created using up to seven PBA-s items [39]. However, multiplying the raw scores together is not considered statistically valid due to the ordinal nature of the severity and frequency data, as multiplication should only be performed on ratio scale data [40]. Additionally, combining the two symptom dimensions contradicts the Rasch model assumption of unidimensionality, which is the concept that all summed items on the scale assess the same underlying construct [32]. Hence, this study aimed to create valid total PBA-s scores by investigating the construct validity of a total PBA-s severity score and separate total PBA-s frequency score.

The total PBA-s severity score

This study revealed that all 11 severity items had disordered response categories, which could either be the result of unclear category labels, too many response categories for a given item, or a combination of these factors. Given that the PBA-s asks for information from the patient, a knowledgeable informant and the clinical impression of the interviewer, combined with detailed additional category descriptions for each item [1], it is unlikely that there is ambiguity

between the majority of category labels. Although, ambiguity may exist semantically between the “slight, questionable” and “mild (present, not a problem)” category labels, demonstrated by the disordered thresholds. Therefore, this study suggests that modifying the PBA-s to a four-point scale improves the construct validity of the total PBA-s severity score, providing a more robust outcome measure for clinical trials.

Further examination of the disordered categories for items referring to rare, but clinically relevant symptoms of suicidal ideation, delusions and hallucinations is needed due to the low category response frequencies observed above category 0 “absent” in this sample [41]. A larger or more advanced HD sample may yield a higher frequency of responses in order to reliably examine the response format of these items.

Although factor analysis in previous literature has identified subscales within the PBA-s [15], the total PBA-s severity score created in this study was deemed to be measuring a unidimensional construct representing the overall severity of behavioural symptoms. The responses for depressed mood were related to the responses for suicidal ideation and the same dependency was found between irritability and aggressive behaviour. Both of these dependencies also make sense at a conceptual level, with regard to the item content. Within this study, a post-hoc grouping of the dependent items was carried out, which accounted for the apparent dependency within this analysis. None of these items were deleted in this study in order to maintain the integrity of the scale, but to further improve the validity of the measurement construct and remove all dependency, a restructuring of the dependent items and their response categories may be considered. For example, creating a single depression item with suicidal ideation integrated into the higher scoring categories.

Also, it was found that females were more likely to endorse item 1 “depressed mood” across all levels of overall severity of behavioural symptoms, which is concordant with reports that the frequency of depression in women is higher in both HD and general populations [42-44]. Although significant DIF (i.e. item bias) existed for depressed mood in this study, the

effects of DIF-cancellation suggested that there was no DIF for gender at the overall scale level, therefore the item was not deleted and the construct validity was uncompromised.

The total PBA-s severity scale was not well targeted, as demonstrated by a lack of overlap between person ability and item difficulty on the person-item threshold distribution map. The threshold map showed clustering of persons at lower levels of the trait (i.e. floor effect) and gaps in the spread of both the items and persons over the range of the construct. The gaps and clustering in persons at lower levels of the trait provide an explanation for the low reliability of the scale as it fails to differentiate persons along the full range of the underlying trait. The mismatch in targeting suggests that additional items need to be incorporated to help discriminate persons at lower levels of the trait. To improve scale targeting and reliability, a review of the original 40-item PBA-HD [1] may provide additional items that are more likely to be endorsed by patients with milder behavioural symptoms, and therefore improve the scaling characteristics among an earlier HD population.

The large floor effect in this study may be expected due to the nature of the study population. The Enroll-HD database consists of a HD population being assessed for motor, cognitive and behavioural symptoms, some of which may report few or no behavioural symptoms. Although the study sample included patients with a wide range of HD severity, better targeting may have been achieved in a more advanced population. However, it is important that the total PBA-s severity score is sensitive to changes in symptom severity early in disease progression, in the hope that early recognition may eventually lead to the potential for prevention.

The total PBA-s frequency score

Attempts to achieve fit to the Rasch model for the PBA-s frequency score were more problematic. Reduction to a universal three-point scale only ordered response categories for 5 out of 11 items and fit to the model remained poor. Some of the disorder may be explained

by low response frequencies for “suicidal ideation”, “delusions” and “hallucinations” items, however, this was not the case for other disordered items. Deletion of items with disordered categories was considered, but due to their clinical relevance they were retained. Again, the disorder may be due to ambiguous category labels or the inclusion of too many response categories. Although the PBA-s frequency labels are detailed, there is a degree of overlap, leading to some categories not being used in the manner intended by the scale developers. For example, category 3 includes the term “most days” and category 4 includes the phrase “almost daily”, which may present significant confusion for both the respondent and the interviewer. Further amendments to the category labels needs to be considered if a total PBA-s frequency score is to be validated as a clinical outcome measure. These results should be interpreted with caution as they may be due to this distinct study population. Further investigation is needed in other samples to support or repute the findings of this preliminary Rasch analysis.

Limitations

Only English speaking countries were included in this study, therefore validation of the total PBA-s severity score in other languages is required to enable generalisation to a truly global HD population. The exclusion of interviews conducted without a knowledgeable informant introduced sampling bias, limiting external validity in a clinical setting where this may not be always possible. Better targeting due to a more advanced HD sample may have been achieved if participants with Juvenile HD were included in the sample. Finally, the usefulness of the total PBA-s severity score without a corresponding total PBA-s frequency score needs to be considered. Restructuring the frequency items could be further explored in order to find a valid approach to successfully assess these two symptom dimensions together. However, the

relevance of frequency data could also be explored, to investigate whether attempts to measure frequency truly add any real information to the severity data alone.

Conclusion

Rasch analysis enabled the psychometric properties of the PBA-s to be examined in more detail than traditional psychometric approaches. This study highlighted important limitations of the PBA-s, primarily the response categories were not being used as intended and there was a lack of overlap between the difficulty of items and the ability of persons in this wide-ranging sample of HD patients. This study recommends that in its current format, the PBA-s should only be used to assess behavioural symptoms individually, as overall symptom scores were not considered to be statistically valid. However, this study found that the PBA-s severity items could be combined to form a valid total score measuring a unidimensional construct, with reduction to a four-point scale, although the mis-targeting suggests that its use as an outcome measure may be more appropriate in a more advanced HD population. Further research concerning the frequency items is needed to determine if similar limitations exist in other samples and to assess the influence of altering response category labels to uncover the potential of a total PBA-s frequency score.

Acknowledgements

This study would like to acknowledge Dr Hugh Rickards and Jenny Callaghan for assistance with project design, Olivia Handley for data collection and Mike Horton for guidance regarding the methodology.

Conflict of Interest

The author has no conflict of interest to report.

References

[1] Craufurd D, Thompson JC, Snowden JS. Behavioral changes in Huntington Disease. Neuropsychiatry Neuropsychol Behav Neurol 2001 Oct-Dec;14(4):219-226.

[2] A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. The Huntington's Disease Collaborative Research Group. Cell 1993 Mar 26;72(6):971-983.

[3] Pringsheim T, Wiltshire K, Day L, Dykeman J, Steeves T, Jette N. The incidence and prevalence of Huntington's disease: a systematic review and meta-analysis. Mov Disord 2012 Aug;27(9):1083-1091.

[4] Evans KR, Sills T, DeBrota DJ, Gelwicks S, Engelhardt N, Santor D. An Item Response analysis of the Hamilton Depression Rating Scale using shared data from two pharmaceutical companies. J Psychiatr Res 2004 May-Jun;38(3):275-284.

[5] Ho AK, Gilbert AS, Mason SL, Goodman AO, Barker RA. Health-related quality of life in Huntington's disease: Which factors matter most? Mov Disord 2009 Mar 15;24(4):574-578.

[6] Hamilton JM, Salmon DP, Corey-Bloom J, Gamst A, Paulsen JS, Jerkins S, et al. Behavioural abnormalities contribute to functional decline in Huntington's disease. J Neurol Neurosurg Psychiatry 2003 Jan;74(1):120-122.

[7] Marder K, Zhao H, Myers RH, Cudkowicz M, Kayson E, Kieburtz K, et al. Rate of functional decline in Huntington's disease. Huntington Study Group. Neurology 2000 Jan 25;54(2):452-458.

[8] Watt DC, Seller A. A clinico-genetic study of psychiatric disorder in Huntington's chorea. Psychol Med 1993;Suppl 23:1-46.

[9] van Duijn E, Kingma EM, van der Mast RC. Psychopathology in verified Huntington's disease gene carriers. J Neuropsychiatry Clin Neurosci 2007 Fall;19(4):441-448.

[10] Carlozzi NE, Miciura A, Migliore N, Dayalu P. Understanding the Outcomes Measures used in Huntington Disease Pharmacological Trials: A Systematic Review. J Huntingtons Dis 2014;3(3):233-252.

[11] Hobart J, Cano S. Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. Health Technol Assess 2009 Feb;13(12):iii, ix-x, 1-177.

[12] Amin L, Rosenbaum P, Barr R, Sung L, Klaassen RJ, Dix DB, et al. Rasch analysis of the PedsQL: an increased understanding of the properties of a rating scale. J Clin Epidemiol 2012 Oct;65(10):1117-1123.

[13] Bech P, Allerup P, Larsen ER, Csillag C, Licht RW. The Hamilton Depression Scale (HAM-D) and the Montgomery-Asberg Depression Scale (MADRS). A psychometric re-analysis of the European genome-based therapeutic drugs for depression study using Rasch analysis. Psychiatry Res 2014 Jul 30;217(3):226-232.

[14] Kingma EM, van Duijn E, Timman R, van der Mast RC, Roos RAC. Behavioural problems in Huntington's disease using the Problem Behaviours Assessment. Gen Hosp Psychiatry 2008 0;30(2):155-161.

[15] Callaghan J, Stopford C, Arran N, Boisse MF, Coleman A, Santos RD, et al. Reliability and factor structure of the Short Problem Behaviors Assessment for Huntington's disease (PBA-s) in the TRACK-HD and REGISTRY studies. J Neuropsychiatry Clin Neurosci 2015 Winter;27(1):59-64.

[16] Zucca A, Lambert SD, Boyes AW, Pallant JF. Rasch analysis of the Mini-Mental Adjustment to Cancer Scale (mini-MAC) among a heterogeneous sample of long-term cancer survivors: a cross-sectional study. Health Qual Life Outcomes 2012 May 20;10:55-7525-10-55.

[17] Rasch G. Probabilistic models for some intelligence attainment tests. Chicago: University of Chicago Press; 1960.

[18] Pallant JF, Tennant A. An introduction to the Rasch measurement model: an example using the Hospital Anxiety and Depression Scale (HADS). Br J Clin Psychol 2007 Mar;46(Pt 1):1-18.

[19] Hobart JC, Cano SJ, Zajicek JP, Thompson AJ. Rating scales as outcome measures for clinical trials in neurology: problems, solutions, and recommendations. Lancet Neurol 2007 Dec;6(12):1094-1105.

[20] Ross CA, Tabrizi SJ. Huntington's disease: from molecular pathogenesis to clinical treatment. Lancet Neurol 2011 Jan;10(1):83-98.

[21] Lerdal A, Kottorp A, Gay CL, Grov EK, Lee KA. Rasch analysis of the Beck Depression Inventory-II in stroke survivors: a cross-sectional study. J Affect Disord 2014 Apr;158:48-52.

[22] Walker FO. Huntington's disease. Lancet 2007 Jan 20;369(9557):218-228.

[23] 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 2011 Apr;26(5):877-884.

[24] Bates G, Harper P, Jones L. Huntington's disease. Oxford: Oxford University Press; 2002. p. 29.

[25] Shoulson I, Fahn S. Huntington disease: clinical care and evaluation. Neurology 1979 Jan;29(1):1-3.

[26] Wheelock VL, Tempkin T, Marder K, Nance M, Myers RH, Zhao H, et al. Predictors of nursing home placement in Huntington disease. Neurology 2003 Mar 25;60(6):998-1001.

[27] Unified Huntington's Disease Rating Scale: reliability and consistency. Huntington Study Group. Mov Disord 1996 Mar;11(2):136-142.

[28] Chatterjee A, Anderson KE, Moskowitz CB, Hauser WA, Marder KS. A comparison of self-report and caregiver assessment of depression, apathy, and irritability in Huntington's disease. J Neuropsychiatry Clin Neurosci 2005 Summer;17(3):378-383.

[29] Andrich D, Lyne A, Sheridan B, Luo G. RUMM2030. Perth, Australia: RUMM Laboratory. 2010.

[30] Andrich D. A rating formulation for ordered response categories. Psychometrika 1978;43:561-573.

[31] Masters G. A Rasch model for partial credit scoring. Psychometrika 1982;47:149-174.

[32] 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 15;57(8):1358-1362.

[33] Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. BMJ 1995 Jan 21;310(6973):170.

[34] Marais I, Andrich D. Effects of varying magnitude and patterns of response dependence in the unidimensional Rasch model. J Appl Meas 2008;9(2):105-124.

[35] Smith EV,Jr. Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. J Appl Meas 2002;3(2):205-231.

[36] Wright B. Rack and Stack: Time 1 vs. Time 2. Rasch Measurement Transactions 2003;17(1):905-906.

[37] Linacre J. Sample size and item calibration stability. Rasch Measurement Transactions 1994;7:328.

[38] Wyse AE. DIF Cancellation in the Rasch Model. J Appl Meas 2013;14(2):118-128.

[39] Tabrizi SJ, Scahill RI, Owen G, Durr A, Leavitt BR, Roos RA, et al. Predictors of phenotypic progression and disease onset in premanifest and early-stage Huntington's disease in the TRACK-HD study: analysis of 36-month observational data. Lancet Neurol 2013 Jul;12(7):637-649.

[40] Field A. Discovering Statistics using IBM SPSS Statistics. 4th ed. Univeristy of Sussex: SAGE Publications Ltd; 2013.

[41] Linacre J. Optimising rating scale effectiveness. Journal of Applied Measurement 2002;3(1):85-106.

[42] Zielonka D, Marinus J, Roos RA, De Michele G, Di Donato S, Putter H, et al. The influence of gender on phenotype and disease progression in patients with Huntington's disease. Parkinsonism Relat Disord 2013 Feb;19(2):192-197.

[43] Aziz NA, Anguelova GV, Marinus J, Lammers GJ, Roos RA. Sleep and circadian rhythm alterations correlate with depression and cognitive impairment in Huntington's disease. Parkinsonism Relat Disord 2010 Jun;16(5):345-350.

[44] Wittchen HU, Jacobi F, Rehm J, Gustavsson A, Svensson M, Jonsson B, et al. The size and burden of mental disorders and other disorders of the brain in Europe 2010. Eur Neuropsychopharmacol 2011 Sep;21(9):655-679.

Tables

Table 1

Structure of the short version of the Problem Behaviours Assessment (PBA-s)

Item	Item description	Severity response categories*
1	Depressed mood	0 = absent
2	Suicidal ideation	1 = slight, questionable
3	Anxiety	2 = mild (present, not a problem)
4	Irritability	3 = moderate (symptom causing problem)
5	Angry or aggressive behaviour	4 = severe (almost intolerable for carer
6	Apathy	**Frequency response categories**
7	Perseverative thinking or behaviour	0 = never/almost never
8	Obsessive-compulsive behaviour	1 = seldom (less than once/week)
9	Paranoid thinking or delusions	2 = sometimes (up to 4 times a week)
10	Hallucinations	3 = frequently (most days/5, 6 or 7 times a week)
11	Disoriented behaviour	4 = daily/almost daily for most (or all) of day

*More detailed scoring criteria and examples exist for the severity response categories in an accompanying manual.

Table 2

Demographics and clinical characteristics of the ENROLL-HD participants (n=517)

Characteristic	Premanifest HD gene-expansion carriers (n=89)	Manifest HD gene-expansion carriers (n=428)		
		TFC I (11-13) (n=104)	TFC II (7-10) (n=186)	TFC III-V (0-6) (n=138)
Age (years)	44.6 (11.5, 20.6-70.8)	54.0 (13.1, 25.0-83.0)	55.7 (12.2, 26.1-84.3)	60.3 (11.6, 27.8-87.7)
Male	33 (37%)	56 (54%)	99 (53%)	62 (45%)
CAG repeat length	42.3 (3.0)	43.1 (2.9)	43.3 (2.8)	43.5 (3.1)
TFC*	12.4 (1.2)	11.9 (0.8)	8.5 (1.1)	4.0 (1.7)

Data are means (SD, range) or number (%).

*TFC: Total Function Capacity (categories I-V), range (0-13).

Table 3

Model fit statistics for PBA-s severity and frequency items

Action	Analysis	Overall model fit	Items fit residual Mean (SD)	Persons fit residual Mean (SD)	PSI	Significant t-tests (%)*
Severity items						
Original	1	χ^2 = 111.28 p = 0.0065	-0.70 (1.28)	-0.22 (0.50)	0.53	4.36
Rescoring all items	2	χ^2 = 84.81 p = 0.059	-0.67 (1.27)	-0.24 (0.53)	0.55	5.05 [95% CI 3.0-7.1%]
Deletion of items 2,9,10	3	χ^2 = 78.19 p = 0.0038	-0.32 (1.25)	-0.25 (0.87)	0.53	4.14
Frequency items						
Original	4	χ^2 = 156.26 p < 0.001	-0.82 (1.25)	-0.19 (0.43)	0.49	4.13
Rescoring all items	5	χ^2 = 142.88 p < 0.001	-0.35 (1.27)	-0.17 (0.47)	0.54	4.82

*CIs are reported when % exceeds 5%. SD: standard deviation; PSI: person separation index (with extremes); χ^2: chi-square; p: probability

Figure Legends

Fig.1. Flow diagram for the inclusion of participants. HD indicates Huntington's disease.

Fig. 2. The category probability curves for item 3 "anxiety" displaying disordered five-point response categories and corrected four-point response categories.

Fig. 3. Item characteristic curve displaying uniform differential item functioning for item 1 "depressed mood" by gender.

Fig. 4. The person-item threshold distribution map for the PBA-s severity items.

Figure 1

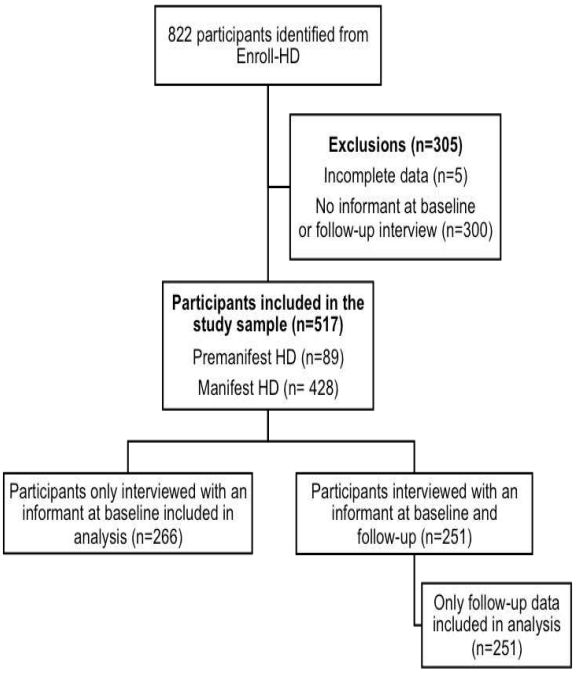

822 participants identified from Enroll-HD
Exclusions (n=305)
Incomplete data (n=5)
No informant at baseline or follow-up interview (n=300)
Participants included in the study sample (n=517)
Premanifest HD (n=89)
Manifest HD (n= 428)
Participants only interviewed with an informant at baseline included in analysis (n=266)
Participants interviewed with an informant at baseline and follow-up (n=251)
Only follow-up data included in analysis (n=251)

Figure 2

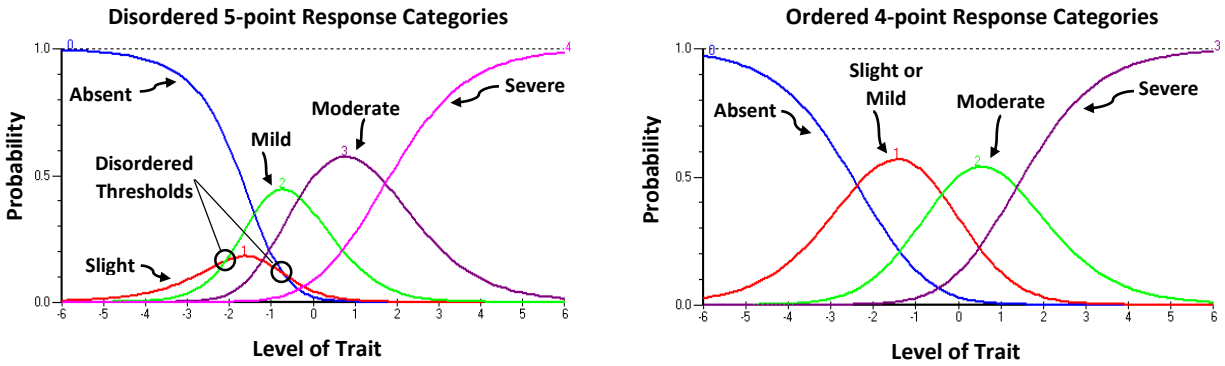

Disordered 5-point Response Categories
Probability
Level of Trait
Absent
Slight
Disordered
Thresholds
Mild
Moderate
Severe
Ordered 4-point Response Categories
Probability
Level of Trait
Absent
Slight or
Mild
Moderate
Severe

Figure 3

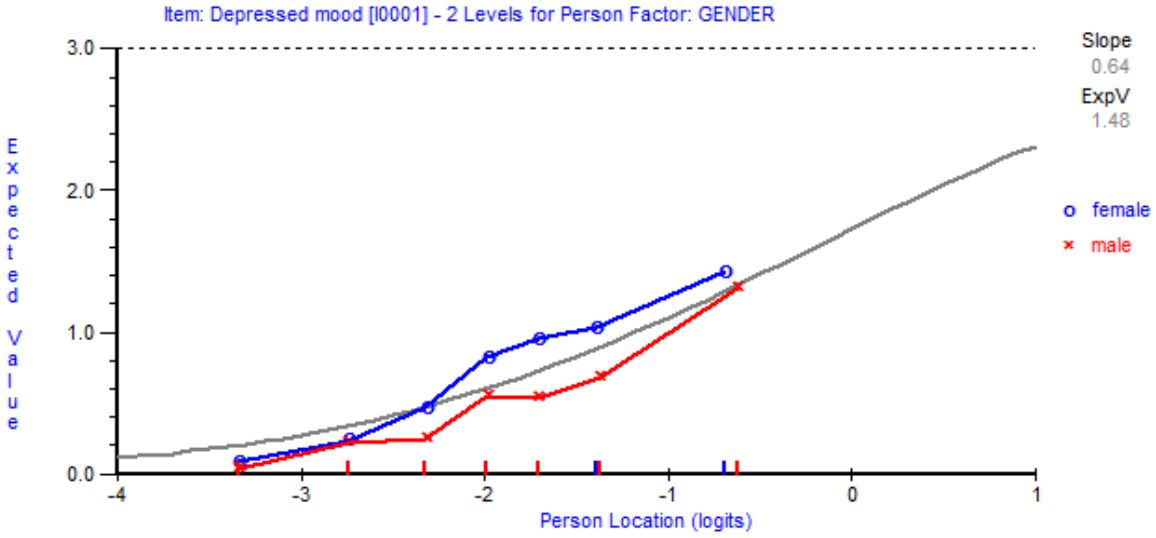
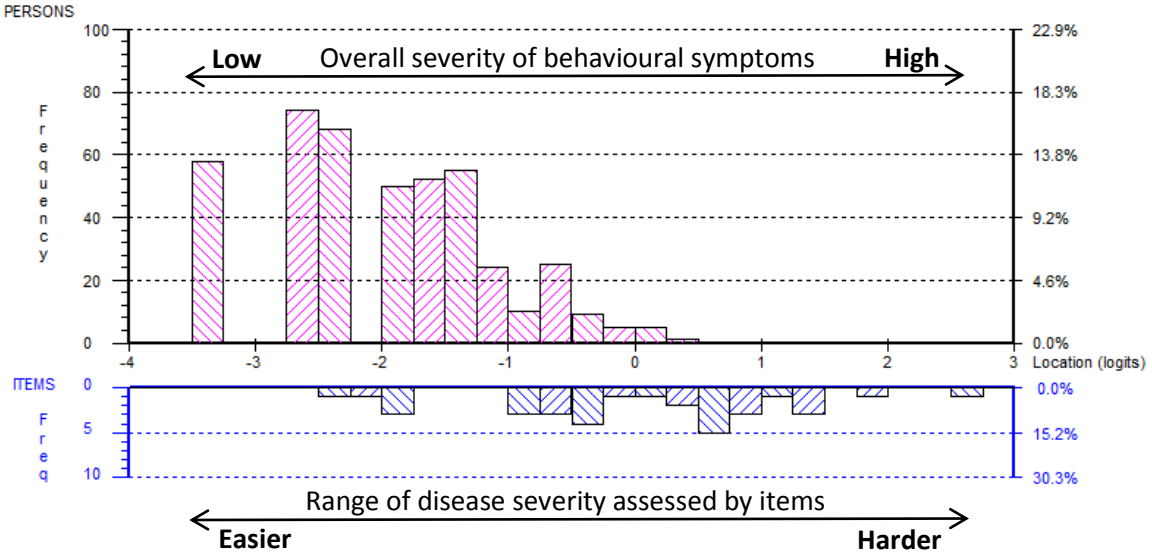

Item: Depressed mood [I0001] - 2 Levels for Person Factor: GENDER
Slope
0.64
ExpV
1.48
female
male
Expected Value
3.0
2.0
1.0
0.0
-4
-3
-2
-1
0
1
Person Location (logits)

Figure 4


PERSONS
Low
Overall severity of behavioural symptoms
High
Frequency
100
80
60
40
20
0
22.9%
18.3%
13.8%
9.2%
4.6%
0.0%
-4
-3
-2
-1
0
1
2
3
Location (logits)
ITEMS
Freq
0
5
10
0.0%
15.2%
30.3%
Range of disease severity assessed by items
Easier
Harder

Appendices

1. The short version of the Problem Behaviours Assessment (PBA-s)

Appendix 1

ENROLL-HD
PROBLEM BEHAVIOURS ASSESSMENT - SHORT (PBA-s)

Participant ☐☐☐-☐☐☐-☐☐☐ **Date** (MM.DD.YYYY) ☐☐.☐☐.☐☐☐☐

All items must be completed. Use U if information is unavailable. Use N if information is not applicable.

	a. Severity:	b. Frequency:	c. Worst:
1. DEPRESSED MOOD:	☐	☐	☐

Suggested prompts:
Start the interview with an open ended question "*Have you noticed any change in your mood since the last visit?*" and then continue with more specific questioning as follows:

- *In the past four weeks, have you been feeling sad? (or blue, or low in spirits?)*
- *Have you found yourself doing something you would ordinarily enjoy and realised you are not having fun?* (Evidence of sad mood from behavioural observation includes sad voice or expression, tearfulness)
- (if yes to either of the above) *Has your mood affected your daily activities?*
- *Does the depressed mood come and go or does it seem always to be there? Is there any change throughout the day? Can you snap out of it if someone tries to cheer you up?*

0 absent
1 questionable
2 low mood is present intermittently but does not interfere with everyday function; rate 2 if participant can easily enjoy amusing activities or visits from friends.
3 participant feels sad much of the time and takes no pleasure from things that he/she usually enjoys, but may still be able to cheer up sometimes with a big effort; rate 3 if low mood has definite effect on participant's lifestyle, e.g. unable to enjoy company of friends or amusing diversions.
4 participant feels sad and utterly miserable all day, takes no pleasure from things that he/she usually enjoys, does not cheer up anytime.
8 unable to assess because condition too advanced (e.g. mute and immobile)

© COPYRIGHT University of Ulm, Faculty of Medicine, EHDN Behavioural Phenotype Working Group (Version 1.4, February 2007). For enquiries, contact permissions@euro-hd.net

2/13

ENROLL-HD
PROBLEM BEHAVIOURS ASSESSMENT - SHORT (PBA-s)

Participant ☐☐☐-☐☐☐-☐☐☐ **Date** (MM.DD.YYYY) ☐☐.☐☐.☐☐☐☐

All items must be completed. Use U if information is unavailable. Use N if information is not applicable.

	a. Severity:	b. Frequency:	c. Worst:
2. SUICIDAL IDEATION:	☐	☐	☐

Suggested prompts:

- *In the past four weeks, have you felt that life was not worth living or that you wouldn't care if you didn't wake in the morning?*
- *have you found yourself thinking that life is not worth living or that you would be better off dead?*
- *Have you thought about harming yourself or even making an attempt at suicide?*
- *Are you planning to hurt yourself or kill yourself? Have you taken any steps towards carrying out your plan?*

0 absent
1 questionable; also rate 1 if participant plans suicide at a later date when disease is more severe but obtains comfort from this as means to retain control of destiny
2 sometimes very pessimistic with fleeting suicidal ideation
3 participant has pervasive and distressing feelings of hopelessness and more prolonged or frequent suicidal ideation, but has not yet acted on this in any way
4 participant has attempted suicide or has made preparations such as saving up tablets or planning ways to avoid discovery when doing it
8 unable to assess because condition too advanced (e.g. mute and immobile)

© COPYRIGHT University of Ulm, Faculty of Medicine,EHDN Behavioural Phenotype Working Group (Version 1.4,February 2007).For enquiries, contact permissions@euro-hd.net

ENROLL-HD
PROBLEM BEHAVIOURS ASSESSMENT - SHORT (PBA-s)

Participant ☐☐☐-☐☐☐-☐☐☐ **Date** (MM.DD.YYYY) ☐☐.☐☐.☐☐☐☐

All items must be completed. Use U if information is unavailable. Use N if information is not applicable.

	a. Severity:	b. Frequency:	c. Worst:
3. ANXIETY:	☐	☐	☐

Suggested prompts:

- *In the past four weeks, have you found yourself getting worried about things? (Evidence of anxiety includes worrying, panic, feeling frightened or fearful for no apparent reason).*
-
- *Have you been worrying a great deal?*
- *What is it like when you worry?*
- *Have you often felt on edge, or keyed up, or mentally strained?*
- *Have you had difficulty in relaxing?*
- *Do your muscles feel tensed up?*
- *When people get anxious or panicky they often feel their heart beating fast or they start shaking or sweating or can't get their breath. Have you had feelings like that?*

0 absent
1 questionable, vague unease (also rate 1 if participant's only worry or anxiety is about prognosis of HD)
2 participant experiences intermittent worry or anxiety, but symptom is not severe enough to cause significant distress or interfere with everyday activities; rate 2 for mild anticipatory anxiety prior to social events or unfamiliar activities e.g. hospital appointments
3 unpleasant anxiety is present much of the time, and has a significant impact on participant's behaviour (e.g. avoids going to places or events associated with provoking anxiety)
4 worry, anxiety or panic are present all the time and have a major impact on participant's lifestyle (e.g. agoraphobia such that participant cannot leave home without an escort)
8 unable to assess because condition too advanced (e.g. mute and immobile)

© COPYRIGHT University of Ulm, Faculty of Medicine,EHDN Behavioural Phenotype Working Group (Version 1.4,February 2007).For enquiries, contact permissions@euro-hd.net

ENROLL-HD
PROBLEM BEHAVIOURS ASSESSMENT - SHORT (PBA-s)

Participant ☐☐☐-☐☐☐-☐☐☐ **Date** (MM.DD.YYYY) ☐☐.☐☐.☐☐☐☐

All items must be completed. Use U if information is unavailable. Use N if information is not applicable.

	a. Severity:	b. Frequency:	c. Worst:
4. IRRITABILITY:	☐	☐	☐

(This item is used to rate the ease with which the participant loses his/her temper, rather than the degree to which self-control is lost once the participant is angry (the latter is rated in the next item). It should also be used to record irritable moods which might have developed into an angry outburst if the carer had not acted with increased tact or discretion)

Suggested prompts:

- *In the past four weeks, have you been irritable, bad-tempered, moody or 'cranky'?*
- *Do you think you get cross more easily than you used to?*
- (if yes to above) *How does this affect the people around you? Do you think they treat you differently when you are like that?*

0 no more irritable than the average person
1 questionable or trivial; within normal limits but worse than he/she used to be
2 definitely more irritable than is reasonable but not to an extent which causes significant problems or distress for other household members; rate 2 if participant appeared to be in a bad mood, but rater considered that participant might have become angry if not treated with tact
3 participant very irritable and looses temper over trivial matters; household members have to be careful what they say and do to avoid problems; rate 3 if participant's appearance and behaviour suggestive of angry mood, such that outbursts would almost certainly have occurred if care had not been taken to placate participant or to keep out of his/her way
4 participant very irritable and looses temper without any obvious reason at all; living with him/her is like walking on eggshells

© COPYRIGHT University of Ulm, Faculty of Medicine,EHDN Behavioural Phenotype Working Group (Version 1.4,February 2007).For enquiries, contact permissions@euro-hd.net

ENROLL-HD
PROBLEM BEHAVIOURS ASSESSMENT - SHORT (PBA-s)

Participant ☐☐☐-☐☐☐-☐☐☐ **Date** (MM.DD.YYYY) ☐☐.☐☐.☐☐☐☐

All items must be completed. Use U if information is unavailable. Use N if information is not applicable.

	a. Severity:	b. Frequency:	c. Worst:
5. ANGRY OR AGGRESSIVE BEHAVIOUR:	☐	☐	☐

Suggested prompts:

- *In the past four weeks, have you had any emotional or angry outbursts?*
- *Have you had times when you lost control of your temper?*
- *Have you hit, shoved or thrown things or expressed your temper in a physical way?*
- *Have you used threats or hostile words?*

0 normal
1 questionable
2 verbal outbursts which are outside socially acceptable limits but do not cause significant problems or distress for other household members; for example, rate 2 if participant becomes angry with self or inanimate objects when confronted with frustrating situations due to disability, such as failure when attempting to rewire a plug.
3 Temper tantrums are severe enough to cause significant distress for other household members and/or practical difficulties caring for the participant; rate 3 when verbal hostility or anger is directed towards another person (e.g. shouting, sarcastic name-calling, use of foul or abusive language). Also rate 3 if there are explicit verbal threats of violence to another person, or behaviour causing a justifiable fear of personal violence (e.g. participant approaches too close, raises fist, mild pushing). Also rate 3 for violence towards property.
4 participant has temper tantrums so severe that relationship with carers is compromised, creating risk that participant will be rejected; rate 4 if there has been any kind of actual physical assault (includes pushing, shoving, hitting, biting, scratching, kicking) or threatening behaviour involving weapons.

© COPYRIGHT University of Ulm, Faculty of Medicine,EHDN Behavioural Phenotype Working Group (Version 1.4,February 2007).For enquiries, contact permissions@euro-hd.net

ENROLL-HD
PROBLEM BEHAVIOURS ASSESSMENT - SHORT (PBA-s)

Participant ☐☐☐-☐☐☐-☐☐☐ **Date** (MM.DD.YYYY) ☐☐.☐☐.☐☐☐☐

All items must be completed. Use U if information is unavailable. Use N if information is not applicable.

	a. Severity:	b. Frequency:	c. Worst:
6. LACK OF INITIATIVE (APATHY):	☐	☐	☐

Suggested prompts:

- *In the past four weeks, have you found that you have lost interest in things that used to be important to you? Are you just as interested as always in trying new things, starting new projects?*
- *Do you have to be pushed to get started on chores that need doing? Do you leave it to friends to take the initiative for organising social activities? Do you sit around a lot doing nothing?*

0 symptom absent
1 questionable
2 Participant no longer tries new things; may need gentle prompting to initiate hobbies or pastimes which he/she usually enjoys; makes less effort to keep up with friends and relatives; tends to put off household tasks which were previously part of normal daily routine and may need gentle prompting to do these things
3 needs quite overt prompting to take part in hobbies or pastimes which he/she used to enjoy, or to carry out routine daily household tasks; makes little or no effort to keep up with friends and leaves it to others to initiate any social contacts; able to take part in (and apparently enjoy) conversation, but tends to follow and is less likely to initiate a change of subject
4 no longer performs any household tasks, even if prompted repeatedly; never initiates activities, and displays no interest in hobbies or pastimes; markedly impoverished speech, rarely initiates new topics of conversation except in relation to own needs; active choices limited to selecting TV programmes to watch, and perhaps switching on or changing channel to do this)
8 unable to assess because condition too advanced (e.g. mute and immobile)

(This item will usually be rated 9 (data missing) in the absence of a reliable informant)

© COPYRIGHT University of Ulm, Faculty of Medicine,EHDN Behavioural Phenotype Working Group (Version 1.4,February 2007).For enquiries, contact permissions@euro-hd.net

ENROLL-HD
PROBLEM BEHAVIOURS ASSESSMENT - SHORT (PBA-s)

Participant ☐☐☐-☐☐☐-☐☐☐ **Date** (MM.DD.YYYY) ☐☐.☐☐.☐☐☐☐

All items must be completed. Use U if information is unavailable. Use N if information is not applicable.

	a. Severity:	b. Frequency:	c. Worst:
7. PERSEVERATIVE THINKING OR BEHAVIOUR:	☐	☐	☐

Suggested prompts:

- *In the past four weeks, have you found yourself getting stuck on certain ideas or actions?*
- *Have your family or friends complained that you are getting obsessed about something, or going on about it more than you should, or doing something over and over again?*

0 symptom absent
1 questionable
2 mild perseverative behaviours or abnormal preoccupations are present but do not interfere with everyday life or cause significant distress for participant or carers; rate 2 if carer reports that participant tends to come out with comments which refer to an earlier topic of conversation, or when rater observes perseverative phenomena during examination (e.g. continues tandem walking after test completed).
3 abnormal preoccupations or repetitive behaviours occupy a significant proportion of participant's attention and cause distress for participant or practical problems for carers; for example, rate 3 if carers report that participant will not let matter drop after an argument, and keeps returning to the same contentious issue all day, or has repetitive behaviours (see below) which cause some interference with everyday care.
4 abnormal preoccupations occupy most of participant's attention for several days at a time, causing major problems or distress for participant and carers, or participant cannot be diverted from repetitive behaviours (pacing, smoking, repeatedly visiting the toilet) which interfere significantly with everyday care.
8 unable to assess because condition too advanced (e.g. mute and immobile)

(This item will usually be rated 9 (data missing) in the absence of a reliable informant)

© COPYRIGHT University of Ulm, Faculty of Medicine,EHDN Behavioural Phenotype Working Group (Version 1.4,February 2007).For enquiries, contact permissions@euro-hd.net

ENROLL-HD
PROBLEM BEHAVIOURS ASSESSMENT - SHORT (PBA-s)

Participant ☐☐☐-☐☐☐-☐☐☐ **Date** (MM.DD.YYYY) ☐☐.☐☐.☐☐☐☐

All items must be completed. Use U if information is unavailable. Use N if information is not applicable.

	a. Severity:	b. Frequency:	c. Worst:
8. OBSESSIVE-COMPULSIVE BEHAVIOURS:	☐	☐	☐

Obsessive-compulsive phenomena are described in the DSM-IV as follows:

- Recurrent or persistent thoughts, impulses or images that are experienced, at some time during the disturbance, as intrusive and inappropriate, and that cause marked anxiety or distress
- The thoughts, impulses or images are not simply excessive worries about real-life problems
- The person attempts to ignore or suppress such thoughts, impulses or images, or to neutralise them with some other thought or action
- The person recognises that the obsessional thoughts, impulses or images are a product of his or her own mind (not imposed from without)
- Repetitive behaviours (e.g. hand-washing, ordering, checking) or mental acts (e.g. praying, counting, repeating words silently) that the person feels driven to perform in response to an obsession, or according to rules that must be applied rigidly
- The behaviours or mental acts are aimed at preventing or reducing distress or preventing some dreaded event or situation; however, these behaviours or mental acts are not connected in any realistic way with what they are designed to neutralise, or are clearly excessive

0 symptom absent
1 questionable or trivial
2 obsessional thoughts or mild compulsive behaviours which do not interfere with everyday life or cause participant significant distress; rate 2 if participant has mild obsessive-compulsive traits such as double checking (a small number of times) that doors are locked or ashtrays empty at night.
3 obsessive-compulsive behaviours are present to a degree which interferes with everyday life or causes significant distress for participant; rate 3 if participant displays mild ritualistic behaviours such as hand-washing, turning lights on and off repetitively or 'evening-up' after touching things by touching with the other hand too.
4 obsessional phenomena cause serious distress, are time consuming (>1 hour/day) or significantly interfere with the person's normal routine, occupational functioning or usual social activities or relationships

© COPYRIGHT University of Ulm, Faculty of Medicine,EHDN Behavioural Phenotype Working Group (Version 1.4,February 2007).For enquiries, contact permissions@euro-hd.net

ENROLL-HD
PROBLEM BEHAVIOURS ASSESSMENT - SHORT (PBA-s)

Participant ☐☐☐-☐☐☐-☐☐☐ **Date** (MM.DD.YYYY) ☐☐.☐☐.☐☐☐☐

All items must be completed. Use U if information is unavailable. Use N if information is not applicable.

	a. Severity:	b. Frequency:	c. Worst:
9. DELUSIONS / PARANOID THINKING:	☐	☐	☐

(Any abnormal beliefs, including unfounded jealous suspicions and accusations of infidelity, should be rated here)

Suggested prompt:

- *I am going to ask you about unusual experiences that people sometimes have. In the past four weeks, has it ever seemed like people are out to get you or perhaps controlling you? Has it seemed like you have special powers or importance, or that books, TV and radio statements are referring to you? Are there any other unusual things you experience that I have not asked you about?*

Suggested <u>additional</u> prompts:

- *Have you felt that people were unduly interested in you or that things were arranged to have special meaning or even that harm might come to you? Can you describe that?*
- *Have there been any other odd or unpleasant experiences of any kind recently?*
- *This would need to be followed by a further exploration of the delusion to establish whether it really is a fixed false belief*

0 symptom absent
1 questionable or trivial
2 overvalued ideas (not amounting to true delusions) are present for some part of the day but do not affect participant's behaviour
3 overvalued ideas are present for much of the day, and participant behaves as if these beliefs were true, although he/she can be persuaded (with difficulty) that he/she is mistaken;
4 delusions: false beliefs, held with unshakeable conviction, which are not shared by other members of participant's social and cultural group and have been present continuously for at least 7 days
8 unable to assess because condition too advanced (e.g. mute and immobile)

(This item will usually be rated 9 (data missing) in the absence of a reliable informant)

© COPYRIGHT University of Ulm, Faculty of Medicine,EHDN Behavioural Phenotype Working Group (Version 1.4,February 2007).For enquiries, contact permissions@euro-hd.net

ENROLL-HD
PROBLEM BEHAVIOURS ASSESSMENT - SHORT (PBA-s)

Participant ☐☐☐-☐☐☐-☐☐☐ **Date** (MM.DD.YYYY) ☐☐.☐☐.☐☐☐☐

All items must be completed. Use U if information is unavailable. Use N if information is not applicable.

	a. Severity:	b. Frequency:	c. Worst:
10. HALLUCINATIONS:	☐	☐	☐

Suggested prompts:

- *In the past four weeks, have you heard things that other people could not hear such as noises or voices of people whispering or talking?*
- *Did you ever have visions (when you were awake) or see things that other people could not see?*
- *How about any other strange sensations in or on your body?*
- *Have you noticed any strange smell or taste that other people seem unable to detect?*

0 symptom absent
1 questionable or trivial
2 participant reports experiencing hallucinations (when asked) but these do not appear to cause any distress or affect participant's behaviour
3 hallucinations which affect participant's behaviour (e.g. looking for source of hidden voices or putting cotton wool in ears) but do not appear to cause much distress
4 participant is clearly distressed by hallucinations and preoccupied with them
8 unable to assess because condition too advanced (e.g. mute and immobile)

Modality of hallucinations: auditory ☐ visual ☐ tactile ☐ olfactory ☐ gustatory ☐

© COPYRIGHT University of Ulm, Faculty of Medicine,EHDN Behavioural Phenotype Working Group (Version 1.4,February 2007).For enquiries, contact permissions@euro-hd.net

ENROLL-HD
PROBLEM BEHAVIOURS ASSESSMENT - SHORT (PBA-s)

Participant ☐☐☐-☐☐☐-☐☐☐ **Date** (MM.DD.YYYY) ☐☐.☐☐.☐☐☐☐

All items must be completed. Use U if information is unavailable. Use N if information is not applicable.

	a. Severity:	b. Frequency:	c. Worst:
11. DISORIENTED BEHAVIOUR:	☐	☐	☐

Suggested prompts:

- *Can you tell me what day of the week it is today? What time is it (morning / afternoon / evening)? Do you know the date? Where are we now?*
- *In the past four weeks, have there been any spells when you were muddled or confused and got these things wrong?*

0 symptom absent
1 questionable or trivial (e.g. participant gets day wrong ± one day, or fails to recognise people when meeting them out of their normal context)
2 participant does not seem to be fully aware of surroundings or the passage of time, but this does not cause significant practical problems
3 evidence of confusion at night (participant appears disoriented in time, place or person to an extent that causes practical problems for carers) but normal during daylight hours
4 participant is confused and disoriented all the time, unaware of time of day / day of week / date and wrongly identifying surroundings or the people around him (e.g. mistakes nursing home for a prison and nursing staff as prison warders) and consequently resists efforts of carers to look after him/her.
8 unable to assess because condition too advanced (e.g. mute and immobile)

© COPYRIGHT University of Ulm, Faculty of Medicine,EHDN Behavioural Phenotype Working Group (Version 1.4,February 2007).For enquiries, contact permissions@euro-hd.net

12/13

ENROLL-HD
PROBLEM BEHAVIOURS ASSESSMENT - SHORT (PBA-s)

Participant ☐☐☐-☐☐☐-☐☐☐ **Date** (MM.DD.YYYY) ☐☐.☐☐.☐☐☐☐

All items must be completed. Use U if information is unavailable. Use N if information is not applicable.

CODES FOR INFORMANT DETAILS:

i). IS INFORMANT A RELATIVE? ☐

1 spouse or partner
2 parent
3 sibling
4 child
5 other relative
6 friend or neighbour
7 professional care worker
8 other
9 no informant - participant came alone

ii). IS INFORMANT A HOUSEHOLD MEMBER? ☐

1 household member (i.e. relative or friend who lives with participant)
2 not a household member but has frequent contact with participant (most days)
3 not a household member and sees participant less than three or four times a week
4 staff of residential care home or hospital

(categories 2 and 3 could apply to family, friends or professional care workers. A paid carer who stays with participant at home for three hours every weekday would be rated as 2).

© COPYRIGHT University of Ulm, Faculty of Medicine,EHDN Behavioural Phenotype Working Group (Version 1.4,February 2007).For enquiries, contact permissions@euro-hd.net

13/13