doi:10.1093/brain/awt048 Brain 2013: 136; 1116–1127 | 1116

BRAIN
A JOURNAL OF NEUROLOGY

The Medical Research Council Prion Disease Rating Scale: a new outcome measure for prion disease therapeutic trials developed and validated using systematic observational studies

Andrew G.B. Thompson,[1,2] Jessica Lowe,[1,2] Zoe Fox,[3] Ana Lukic,[1,2] Marie-Claire Porter,[1,2] Liz Ford,[2] Michele Gorham,[2] Gosala S. Gopalakrishnan,[2] Peter Rudge,[1,2] A. Sarah Walker,[4] John Collinge[1,2] and Simon Mead[1,2]

1 MRC Prion Unit, Department of Neurodegenerative Disease, University College London (UCL) Institute of Neurology, UCLH NHS Trust, Queen Square, London, UK
2 National Prion Clinic, National Hospital for Neurology and Neurosurgery, UCLH NHS Trust, Queen Square, London, UK
3 Joint Research Office Biostatistics Unit, UCL, and Education Unit, UCL Institute of Neurology, London, UK
4 MRC Clinical Trials Unit, London, UK

Correspondence to: Professor John Collinge, MD, FRS, FRCP, MRC Prion Unit, Department of Neurodegenerative Disease, UCL Institute of Neurology, Queen Square, London, WC1N 3BG, UK
E-mail: j.collinge@prion.ucl.ac.uk

Progress in therapeutics for rare disorders like prion disease is impeded by the lack of validated outcome measures and a paucity of natural history data derived from prospective observational studies. The first analysis of the UK National Prion Monitoring Cohort involved 1337 scheduled clinical assessments and 479 telephone assessments in 437 participants over 373 patient-years of follow-up. Scale development has included semi-quantitative and qualitative carer interviews, item response modelling (Rasch analysis), inter-rater reliability testing, construct analysis and correlation with several existing scales. The proposed 20-point Medical Research Council Prion Disease Rating Scale assesses domains of cognitive function, speech, mobility, personal care/feeding and continence, according to their relative importance documented by carer interviews. It is quick and simple to administer, and has been validated for use by doctors and nurses and for use over the telephone, allowing for frequent assessments that capture the rapid change typical of these diseases. The Medical Research Council Scale correlates highly with widely used cognitive and single item scales, but has substantial advantages over these including minimal floor effects. Three clear patterns of decline were observed using the scale: fast linear decline, slow linear decline (usually inherited prion disease) and in some patients, decline followed by a prolonged preterminal plateau at very low functional levels. Rates of decline and progress through milestones measured using the scale vary between sporadic, acquired and inherited prion diseases following clinical expectations. We have developed and validated a new functionally-oriented outcome measure and propose that future clinical trials in prion disease should collect data compatible with this scale, to allow for combined and comparative analyses. Such approaches may be advantageous in orphan conditions, where single studies of feasible duration will often struggle to achieve statistical power.

Received October 31, 2012. Revised December 18, 2012. Accepted January 31, 2013

Keywords: prion; cohort; rating scale; outcome measure; CJD

Abbreviations: CJD = Creutzfeldt-Jakob disease; MRC = Medical Research Council

Introduction

The prion diseases are a group of rare neurodegenerative conditions for which no proven disease-modifying treatment is available. They are characterized by templated misfolding of the normal cellular prion protein (PrP^{c}) into abnormal disease-associated forms (generally referred to as PrP^{Sc}) but they may have genetic, idiopathic or acquired aetiologies. There is remarkable clinical heterogeneity both within and between these different aetiological disease types (Collinge, 2001).

The transmissibility of prion disease both within and between mammalian species, leading to fatal neurodegeneration that faithfully reproduces the clinicopathological features of the human disease, provides a uniquely robust opportunity for laboratory validation of experimental therapeutics for a human neurodegenerative disease. As a result of these animal models, and rapidly advancing understanding of the molecular pathogenesis of prion disease, development of putative therapeutic agents for use in human trials (both small molecules and monoclonal antibodies that bind to PrP^{c}) has now reached an advanced stage (Mallucci *et al.,* 2002, 2003, 2007; White *et al.,* 2003; Nicoll and Collinge, 2009; Nicoll *et al.,* 2010; Klohn *et al.,* 2012). However, if the promise of these agents is to be realized, it is vital that trial methodology is optimized and a number of challenges specific to prion disease trials are overcome. Even a highly effective therapeutic agent may fail to produce conclusive results without an optimal trial design.

The Medical Research Council (MRC) PRION-1 trial demonstrated that a large clinical trial in a single country is feasible, but was limited by the lack of a validated measure of clinical progression (Collinge *et al.,* 2009). Anticipating this issue, the trial protocol included a variety of existing rating scales designed to probe neurological, cognitive, psychiatric and general functional status. Analysis of the performance of eight of these scales in PRION-1 in terms of validity, practicality and statistical power in simulated clinical trials supported the use of functionally-orientated measures relative to global, neurological, cognitive or psychiatric scales (Mead *et al.,* 2011). However, a number of unresolved issues remained: no single scale captured progression across the full range of functional physical and cognitive domains affected by any of the individual categories of prion disease; patients could not be visited frequently enough to capture the very rapid decline that is typical of Creutzfeldt-Jakob disease (CJD); floor effects (large numbers of subjects with the worst possible score) were observed in all scales except the Glasgow Coma Score; and the patient sample was too small to allow reliable analysis of important aetiological or severity-stratified subgroups (Mead *et al.,* 2009).

With the aim of developing a single, functionally-orientated and validated outcome measure, tailored to the particular demands of a prion disease clinical trial, we modified and combined elements of three rating scales that showed relative strengths in the PRION-1 analysis and are well established in other neurological settings: the Modified Barthel Activities of Daily Living Index (Barthel) (Mahoney and Barthel, 1965; Collin *et al.,* 1988), the Clinical Dementia Rating Sum of Boxes (Morris, 1993), and the Glasgow Coma Score (Teasdale and Jennett, 1974). We carried out semi-quantitative interviews to assess patients' relatives and carers on which manifestations of prion disease are of greatest concern to them, and ensured that these were reflected in the domains assessed by our scale. We assessed reliability of use over the telephone (allowing much higher assessment frequency), and used Rasch analysis (Hobart *et al.,* 2007) to refine an outcome measure that had the most favourable statistical properties when used in this patient group. Finally, we illustrate the natural history of symptomatic patients in the National Prion Monitoring Cohort study and/or PRION-1 trial using the MRC Prion Disease Rating Scale, and performance relative to existing scales.

Materials and methods

Patient referral, clinical diagnosis and enrolment

A national referral system for prion diseases was set up in the UK in 2004. UK neurologists were asked by the Chief Medical Officer to refer all patients with suspected prion disease jointly to the National CJD Research and Surveillance Unit (Edinburgh, UK) and to the NHS National Prion Clinic (London, UK). This enables epidemiological surveillance, provision of specialist clinical care and also participation in clinical research, including the PRION-1 trial (2001–2007) (Collinge *et al.,* 2009) and the National Prion Monitoring Cohort study (subsequently 'Cohort study', 2008–2012). Collinge *et al.* (2009) provides details of enrolment into the PRION-1 trial, which were similar to those described for the Cohort study below.

The Cohort study began in October 2008, and aimed to enrol all symptomatic patients with prion disease in the UK thereafter. This includes all cases of probable or definite prion disease (sporadic CJD, variant CJD, iatrogenic CJD, and inherited prion disease). Also eligible for enrolment are asymptomatic individuals known to be at risk of inherited prion disease (tested asymptomatic gene mutation carriers or untested first degree relatives of those with a confirmed pathogenic *PRNP* mutation), or variant CJD (recipients of implicated whole or leucodepleted blood transfusion notified by the Health Protection Agency). A small group of healthy control subjects were also recruited (friends or relatives without pathogenic *PRNP* mutations or other known risk factors). 'At risk' individuals and healthy control subjects did not contribute to the development of rating scales.

Enrolment by National Prion Clinic staff took place at hospitals, nursing homes, hospices and patients' homes around the UK. Diagnoses and eligibility were reviewed by senior National Prion Clinic clinicians (J.C., S.M. and/or P.R.) within a week of enrolment, or prior to enrolment if there was clinical uncertainty. Probable sporadic CJD diagnosis was made according to World Health Organization criteria, with the addition of brain MRI as a supportive investigation following recent recommendations (pathological signal change on FLAIR or diffusion weighted sequences in the basal ganglia, thalamus

and/or ⩾2 cortical regions) (Zerr *et al.,* 2009). Patients not meeting criteria for sporadic CJD were enrolled if this was thought to be the most likely diagnosis at reveiw by a panel of senior NPC clinicians. Probable variant CJD diagnosis was made according to World Health Organization criteria (WHO, 2001). Inherited prion disease was diagnosed by gene test. Iatrogenic CJD was diagnosed using sporadic CJD criteria with a relevant history of exposure. If patients died during the study and underwent post-mortem examination, or had relevant tissue biopsy (brain or tonsil) during life, the pathological results were used to confirm diagnosis.

A further 26 patients were recruited but not included in study participant totals or any analysis because an alternative diagnosis became more likely than prion disease during follow-up, due to their clinical course (e.g. persistent improvement) and/or clinical investigation results (e.g. presence of serum voltage-gated potassium channel complex antibodies, post-mortem results).

Consent and ethics

Informed consent was obtained directly from study participants if possible, or otherwise from relatives, carers or Independent Mental Capacity Advocates as appropriate. Ethical approval was obtained from the Scotland A Research Ethics Committee (Cohort) or the Eastern Research Ethics Committee (PRION-1).

Stratification and assessment schedule

Participants were stratified at enrolment to the Cohort according to their likely rate of disease progression based on diagnosis:

Stratum 1: Symptomatic patients with sporadic CJD, variant CJD, iatrogenic CJD and forms of inherited prion disease likely to have rapid clinical progression (due to 4-octapeptide repeat insertion, E200K, D178N, E211Q or V210I mutations); Stratum 2: Symptomatic patients with inherited prion disease likely to have slow progression (those with 5 and 6-octapeptide repeat insertion, P102L, P105L, Q212P, A117V or Y163X mutations); and Stratum 3: At-risk and healthy control individuals.

Patients in Stratum 1 had face-to-face follow-up assessments initially every 6–8 weeks. If clinical progression proved to be slower than the expected significant deterioration over this interval (i.e. minimal or no change by overall clinical impression over 6–8 weeks), study physicians could decide to reduce follow-up frequency for subsequent assessments (up to a maximum interval of 24 weeks). Patients in Strata 2 and 3 had follow-up assessments every 6 to 12 months. The following rating scales were administered at study assessments: Modified Barthel Activities of Daily Living index (Mahoney and Barthel, 1965; Collin *et al.,* 1988), Clinical Dementia Rating Sum of Boxes (Morris, 1993), Glasgow Coma Score (Teasdale and Jennett, 1974), Rankin (Rankin, 1957), and Mini-Mental State Examination (Folstein *et al.,* 1975). In September 2009 an initial scale was also introduced, as described below. In addition, a systematic clinical history, neurological examination and a short neuropsychological test battery were done at all assessments, unless precluded by the level of disease severity or patient fatigue. From May 2010 patients in Stratum 1 also had telephone assessments every 1 to 2 weeks between follow-up assessments. These were discontinued if there was minimal change between consecutive assessments (at worst possible score).

Survey of carer priorities

The Barthel Activities of Daily Living Caregiver Interview Summary form was used to gather semiquantitative data from relatives/carers on the aspects of prion disease of greatest importance to them and the patients in their daily lives. Up to four symptoms or impairments of greatest concern to them at the time of assessment were ranked in order of importance. This was completed at 69 assessments, by carers of patients affected by different prion disease types. These data were supplemented by discourse analysis of 10 in-depth structured interviews with carers.

Development pathway for novel rating scale

Building on and extending the analysis of the pre-existing scales in PRION-1 (Mead *et al.,* 2011), an iterative approach was taken to develop and refine a novel scale for use in prion disease. An initial version took into account the relative performance and limitations of different pre-existing scales and their subcomponents in the PRION-1 analysis, the reported priorities of patients and carers, the need to include a range of both physical and cognitive domains, and the pooled clinical experience of the National Prion Clinic medical, nursing and neuropsychology staff. This scale version was then administered at all Cohort assessments alongside the range of scales above (from September 2009), as well as over the telephone (from May 2010).

The performance of this scale version was assessed with respect to ease of use, ability to capture clinically evident decline, inter-rater reliability between treating clinicians and nurses, reliability of use over the telephone, floor and ceiling effects. As it consisted of a combination of slightly modified subcomponents from existing scales used in PRION-1 and throughout the Cohort, a close approximation of the scale could be calculated for all assessments carried out prior to its introduction, allowing a larger number of patients and assessments to be included in sensitivity/additional analysis.

Rasch analysis

Rasch analysis was performed on the first two initial scale version assessments post September 2009 per individual, to ensure that the analysis was not unduly affected by a small minority of patients with many assessments, and to permit testing the function of the scale over time. Rasch analysis was used to see whether the initial scale fitted this model using Rasch Unidimensional Measurement Model 2030 software. The partial credit variant of the polytomous model was chosen to reflect multiple successively ordered response categories of varying difficulty (Andrich *et al.,* 2008). We took the following recommended (Tennant and Conaghan, 2007) iterative steps to optimize the fit of our scale to the Rasch model: (i) exploring item–person interactions (to examine the degree to which the Guttman pattern was achieved) and item–trait interactions using χ^2-based fit statistics; (ii) rescoring items that demonstrated disordered thresholds (i.e. an item's scoring categories do not progress in a logical order); (iii) removing the most poorly fitting items, or combining items into super-items where appropriate; (iv) examining local dependencies between items that could confound the assumption of unidimensionality of the scale, and dropping items when misfit was still apparent; (v) investigating differential item functioning for gender, age, time of assessment (first versus second assessment), assessor (doctor versus nurse administering scale) and mode of assessment (telephone versus face-to-face). We also analysed a random sample of later follow-up assessments to evaluate whether there was differential item functioning for the selection of the follow-up visit to include in the main analysis; (vi) examining local dependencies using residual correlation matrices, and exploring unidimensionality of the scale using principal components analysis to select the highest positive and highest negative items

followed by *post hoc t*-tests on person locations determined by these items; and (vii) reviewing summary fit statistics after all modifications to the scale were made. Cronbach's Alpha statistic was used to confirm reliability of the fit statistics. The final scale, termed the MRC Prion Disease Rating Scale (or MRC Scale) was selected when satisfactory fit was achieved overall, no items showed poor fit and the scale was shown to be compatible with unidimensionality. To increase numbers, a sensitivity analysis also used scores derived from the individual component scales (Barthel, Clinical Dementia Rating Sum of Boxes, and Glasgow Coma Score) prior to September 2009 (Tools was scored as missing) for individuals who did not have two scores post-September 2009.

Results

Patients

Four hundred and thirty-seven participants consecutively enrolled into the Cohort and/or PRION-1 studies (up to April 2012) were included in the analysis; 240 patients with sporadic CJD, 25 with variant CJD, 12 with iatrogenic CJD, and 81 with symptomatic inherited prion disease (three, seven and 19 patients with 4, 5 and 6 octapeptide repeat insertion, respectively, seven with A117V mutation, four with D178N, eight with E200K, 26 with P102L, one patient each with E211Q, P105L, Q212P and V210I mutation and three with Y163X) together with 34 individuals at risk of inherited prion disease, 10 at risk of variant CJD, and 35 healthy control subjects. Three hundred and eleven patients died during the study, 192 (62%) of these underwent post-mortem examination, which confirmed the diagnosis of prion disease in all cases. Eighty-nine (20%) had diagnosis confirmed in life by gene test (all 81 inherited prion disease) or tissue biopsy ($n = 8$) some also with post-mortem examination. One hundred and fifty-five (35%) were diagnosed using clinical and investigation findings alone.

Figure 1 illustrates enrolment, stratification, follow-up and drop-out from the studies. Ninety-seven per cent of symptomatic patients judged eligible for the Cohort study at the initial National Prion Clinic assessment were enrolled, suggesting that the study is highly effective at capturing recognized prion disease in the UK. Less than 1.5% withdrew from the studies. Baseline characteristics are shown in Table 1.

Scale development

Table 2 summarizes the scale development process. Ranked symptoms and impairments reported by caregivers were grouped into functional domains. The most frequently recorded were within

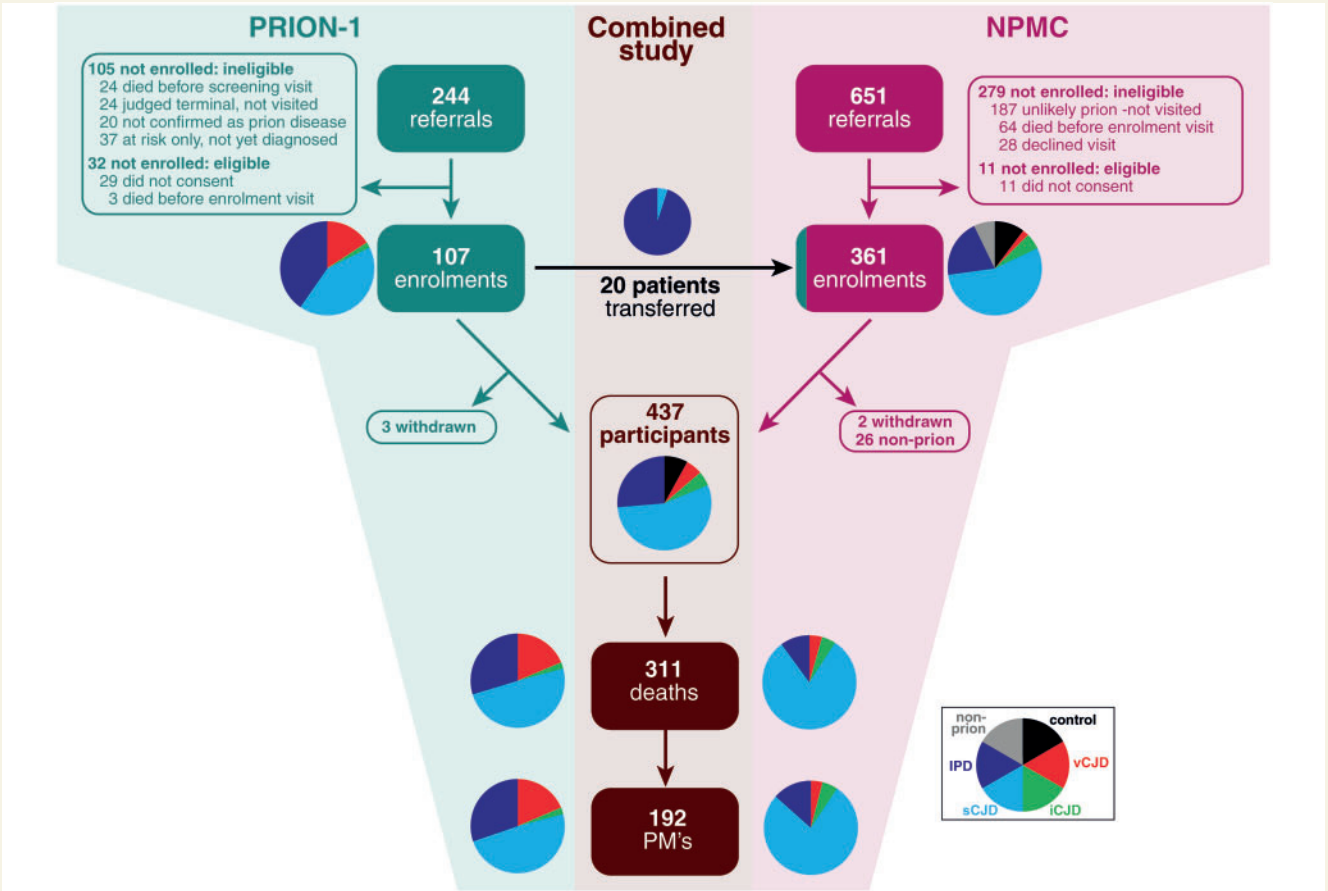

Figure 1 Study profile. iCJD = iatrogenic CJD; IPD = inherited prion disease; NPMC = National Prion Monitoring Cohort; PM = post-mortem; sCJD = sporadic CJD; vCJD = variant CJD.

Table 1 Baseline characteristics at enrolment

	All patients	sCJD	IPD	vCJD	iCJD	At risk	Control
Enrolled	**437**	**240**	**81**	**25**	**12**	**44**	**35**
Median age (years; range)	61 (14–92)	67 (40–87)	48 (26–92)	30 (14–63)	42 (27–51)	42 (20–92)	48 (23–75)
Gender (M/F)	207/230	105/135	39/42	17/8	10/2	19/25	17/18
Median time (months) since first symptoms (IQR)	6 (3–13)	4 (2–8)	26 (7–63)	8 (5–10)	8 (5–14)	N/A	N/A
Rankin							
Number assessed	**430**	**238**	**79**	**23**	**11**	**44**	**35**
Asymptomatic (0)	75	0	0	0	0	41	34
No or slight symptoms (1/2)	29	4	18	2	3	1	1
Moderate disabilty (3)	51	18	26	6	1	0	0
Moderate to severe disability (4)	89	56	17	12	3	1	0
Severe disability (5)	186	160	18	3	4	1	0
Barthel index							
Number assessed	**424**	**238**	**76**	**20**	**12**	**43**	**35**
Median (IQR)	4 (0–19)	0.5 (0–3.5)	14.5 (2–19.5)	10 (5.6–16.1)	6.3 (2–17)	20	20
MMSE							
Number assessed	**385**	**205**	**72**	**19**	**11**	**43**	**35**
Median (IQR)	12 (0–27)	0 (0–10)	18 (5–24.8)	16 (12–21.5)	15 (1.5–27)	30 (29–30)	30
CDR							
Number assessed	**353**	**178**	**69**	**21**	**9**	**42**	**34**
Median (IQR)	11 (2–18)	18 (12–18)	8 (3–15)	11 (5.5–16)	10 (4–18)	0	0
GCS							
Number assessed	**399**	**232**	**64**	**20**	**10**	**38**	**35**
Median (IQR)	14 (10–15)	11 (9–14)	15 (11–15)	15 (13.8–15)	14 (13.3–15)	15	15
MRC Scale							
Number assessed	**388**	**212**	**67**	**18**	**12**	**44**	**35**
Median (IQR)	9 (2–19)	3 (1–8.8)	16 (5.5–19)	13 (9–17.5)	11.5 (5–17.3)	20	20

Interquartile ranges (IQR) are shown. 'At risk' and 'Control' groups did not contribute directly to development of the novel scale.
CDR = Clinical Dementia Rating Sum of Boxes; GCS = Glasgow Coma Score; iCJD = iatrogenic CJD; IPD = inherited prion disease; MMSE = Mini-Mental State Examination; sCJD = sporadic CJD; vCJD = variant CJD.

domains of mobility (55/236 ranked items, 23%), personal care/continence (30/236, 13%), communication/speech (29/236, 12%), behaviour/hallucinations (29/236, 12%), eating/swallowing (27/236, 11%) and cognition/memory/navigation (21/236, 9%). Symptoms or impairments from other domains were reported in $< 6\%$ of responses. These were the key domains for representation in our outcome measure.

The initial scale, based on that proposed in Mead *et al.* (2011), consisted of slightly modified versions of all subcomponents of the Modified Barthel Activities of Daily Living Index, the memory, orientation and judgement/problem-solving subcomponents of the Clinical Dementia Rating Sum of Boxes, the best verbal response subcomponent of the Glasgow Coma Score and a novel subcomponent assessing ability to use tools (15 items, 35 thresholds) (Mead *et al.*, 2011). It was designed for completion based on a brief interview with a closely-involved relative or carer as inability of patients to participate was a major cause of poor completion rates for some scales in PRION-1. Compared with their parent scales, minor modifications were made to a number of subcomponents at this stage to make them more easily applicable to patients with prion disease e.g. to account for severe expressive dysphasia/mutism, and to add additional intermediate response categories at severe levels of impairment, aiming to improve discrimination and reduce floor effects.

The initial scale was completed on a total of 977 occasions, in 266 patients. This consisted of 498 face-to-face assessments and 479 telephone assessments. The scale proved to be simple and easy to use, being completed in less than 5 min. In addition to this, an approximation of the initial scale could be calculated from Barthel, Clinical Dementia Rating Sum of Boxes and Glasgow Coma Score scale data for a further 839 assessments pre-September 2009.

The initial scale was acquired over the telephone by a doctor and a specialist nurse within 24 h of each other on 50 occasions, over the telephone by a nurse within 24 h of a face-to-face assessment by a doctor on two occasions, and by both doctor and nurse attending the same visit on two occasions. In all cases each assessor was blinded to the other's score. Agreement across all paired assessments was good or excellent (Cohen's kappa > 0.6) for all but three items: dressing, which was subsequently dropped from the scale; best verbal response, which subsequently underwent item threshold rescoring; and orientation, which was subsequently merged with memory.

The fit to the Rasch model of the 15-item initial scale was first assessed in all individuals across all symptomatic disease types, and demonstrated poor fit in this heterogeneous population ($\chi^2 = 166.16$, df = 30, $P < 0.0001$). As Rasch analysis relies on the assumption that there is a single construct being measured by all items in a scale, including patients with different disease types that are known to vary widely in their typical clinical progression, it is

Table 2 Summary of scale development

Initial scale		Carer priorities and clinical validity			Inter-rater reliability		Rasch Analysis		MRC Prion Disease Rating Scale		
Item	Derived from		Problems identified	Modifications	Kappa ($n = 54$)	Modifications	Problems identified	Modifications	Included?	Recalculated Kappa ($n = 54$)	Maximum Score (Total 20)
Bowel function	Barthel	Continence	–	–	0.74	–	Fit improved by rescoring	Adjacent levels collapsed	Yes	0.78	1
Bladder function	Barthel	Continence	–	–	0.71	–	Fit improved by rescoring	Adjacent levels collapsed	Yes	0.89	1
Grooming	Barthel	Personal care	Personal care over-represented relative to other domains	–	0.67	–	Item dependency with bathing	Item dropped	No	–	–
Toilet use	Barthel	Personal care	–	–	0.81	–	–		Yes	0.81	2
Feeding	Barthel	Personal care	Floor effect	Extra level added	0.70	–	Fit improved by rescoring	Adjacent levels collapsed	Yes	0.83	2
Transfers	Barthel	Mobility	–	–	0.73	–	Item dependency with Mobility	Items merged	Yes	0.67	2
Mobility	Barthel	Mobility	'Wheelchair independent' very rarely applicable	Rewording	0.81	–	Levels disordered				
Dressing	Barthel	Personal care	–	–	0.39	Item dropped	–	–	No	–	–
Stairs	Barthel	Mobility	–	–	0.82	–	–	–	Yes	0.82	2
Bathing	Barthel	Personal care	Floor effect	Extra level added	0.66	–	Fit improved by rescoring	Adjacent levels collapsed	Yes	0.73	1
Best verbal response	GCS	Speech	–	–	0.85	–	–	–	Yes	0.85	4
Use of tools	Novel item	Cognition	–	–	0.72	–	–	–	Yes	0.72	1
Memory	CDR	Cognition	Expressive dysphasia typically obscures verbal memory	Rewording, to account for dysphasia	0.64	–	Fit improved by rescoring	Adjacent levels collapsed; item merged	Yes	0.70	3
Orientation	CDR	Cognition	Floor effect	Extra level added	0.80	Item merged	Item dependency with memory/ language				
Judgement/ problem solving	CDR	Cognition	–	–	0.74	–	Fit improved by rescoring	Adjacent levels collapsed	Yes	0.82	1

plausibly impossible to establish a fit to the Rasch model using the whole data set. As our primary objective was to develop an outcome measure tailored to sporadic CJD, all other disease groups were excluded and the analysis repeated using 205 records from 132 patients with sporadic CJD (all of those in whom the initial scale had been administered). Fit improved somewhat, but was still unsatisfactory ($\chi^2 = 77.01$, df = 30, $P < 0.0001$) so further exploratory analyses were performed to identify the reasons for this.

The 15 individual items were examined for threshold ordering, individual item fit, differential item functioning and local item dependencies, and changes made to address these issues as recommended with consequent improvements in fit (e.g. rescoring of disordered items, dropping of most poorly fitting item) (Tennant and Conaghan, 2007). These changes are summarized in Table 2. Item dependencies were identified between grooming and bathing; transfers and mobility; toilet and stairs; bathing, orientation and memory. We created two 'super-items' by combining pairs of heavily dependent items in a clinically meaningful way: transfers and mobility; and memory and orientation. The structure and scoring of these super-items can be seen in Table 3. Once item

Table 3 Final MRC Prion Disease Rating Scale

Item	Category criteria	Score
Bowel function	At least one episode of incontinence in last 7 days	0
	Continent for last 7 days	1
Bladder function	Always incontinent or catheterized	0
	Continent or occasional accidents	1
Toilet use	Fully dependent	0
	Needs some help	1
	Independent	2
Bathing	Fully dependent or needs some help	0
	Independent	1
Feeding	Unable or NG/PEG/RIG fed (takes nothing by mouth)	0
	Needs help but can swallow (even if unsafe)	1
	Independent	2
Transfers and mobility	Bedbound, unable to sit	0
	Can sit, but cannot mobilize or transfer without help (from person or walking aid)	1
	Can transfer or mobilize independently or both	2
Stairs	Unable	0
	Needs help	1
	Independent	2
Best verbal response	Mute	0
	Incomprehensible sounds	1
	Single words	2
	Sentences, but difficulty in finding words, uses incorrect words or is often disoriented/confused	3
	Normal conversation	4
Memory and orientation to surroundings	Shows no awareness of surroundings or any evidence of memory	0
	Evidence of retaining some highly learned material (e.g. recognizing familiar people) or awareness of surroundings but no evidence of acquiring new material	1
	Able to retain some new information but memory consistently impaired	2
	Memory normal or some impairment off and on	3
Judgement and problem solving	Unable to show any judgement or problem-solving	0
	Able to show some judgement or problem-solving, even if this is severely impaired	1
Use of tools	Unable to use any tools or objects	0
	Able to use some tools or objects, with help if necessary	1

NG = nasogastric; PEG = percutaneous endoscopic gastrostomy; RIG = radiologically inserted gastrostomy.

rescoring, appropriate ordering and local dependencies were addressed, items were dropped if fit was still not achieved.

A revised scale including all of these modifications, termed the MRC Scale, demonstrated a good fit ($\chi^2 = 25.19$, df = 22, $P = 0.29$). The mean fit residual for all 11 items in this scale was -0.498 with a standard deviation (SD) of 0.993, illustrating a good fit of the items to the model. Cronbach's alpha was 0.91, implying a high level of confidence in the reliability of the assessment of fit, and the Person Separation Index was 0.90. No individual item had a fit residual of magnitude >2. Examination for all items across all person factors confirmed the data to be free of differential item functioning when Bonferroni probability adjustment for multiple testing was applied, with gender, age quartile, assessor (doctor or nurse), assessment mode (face-to-face or telephone) included as person factors.

Principal components analysis demonstrated that the greatest variance in the data existed between the most positively loaded items; those assessing speech and cognitive functions (speech, memory/orientation, judgement and use of tools) and the most negatively loaded items which assessed some aspects of mobility, personal care and continence (toilet, bowels, bladder, stairs, mobility/transfers). Feeding and bathing had no strong loading. Comparing the person locations from positively loaded items with a factor loading >0.4 (speech and memory/orientation) to negatively loaded items with a factor loading < -0.4 (toilet, stairs, bladder, bowels and transfers/mobility) produced significant results at the 1% level in 9/205 (4.4%) individuals and at the 5% level in 16/205 (7.8%) individuals. Overall, our analyses of multidimensionality suggest that differential progression in speech/cognitive and mobility/personal care/continence domains is the second most important dimension (after the Rasch model) but did not grossly compromise the unidimensionality of the scale and the construct of sporadic CJD disease progression.

Including earlier data derived from the original component scales (Barthel, Clinical Dementia Rating Sum of Boxes, and Glasgow Coma Score) along with the data collected using the initial scale allowed a total of 380 scales in 239 patients with sporadic CJD to be analysed. The overall fit to the Rasch model was less good but remained acceptable ($\chi^2 = 33.75$, df = 22, $P = 0.052$), suggesting that the modified items introduced for the initial scale were performing better than pre-existing scale components. These item modifications were therefore retained in the final scale. Testing the MRC Scale's fit in all Stratum 1, rapidly progressive patients, including a rapidly progressive subset of inherited prion disease along with sporadic CJD, also demonstrated acceptable fit to the model ($\chi^2 = 30.14$, df = 22, $P = 0.11$). Unsurprisingly, testing fit for patients with inherited prion disease alone and for all symptomatic patients (all disease groups including Stratum 2) continued to result in a poor fit due to the heterogeneity of clinical syndromes that comprise inherited prion disease (such as the distinction between the late cognitive features of Gerstmann-Sträussler-Scheinker syndrome, and early predominant cognitive decline in 6-octapeptide repeat insertion prion disease).

The final format of the rating scale is shown in Table 3. The inter-rater reliability in administration between doctors and nurses of the final scale as a whole was excellent (interclass correlation coefficient = 0.96). Figure 2 shows the correlation of the MRC

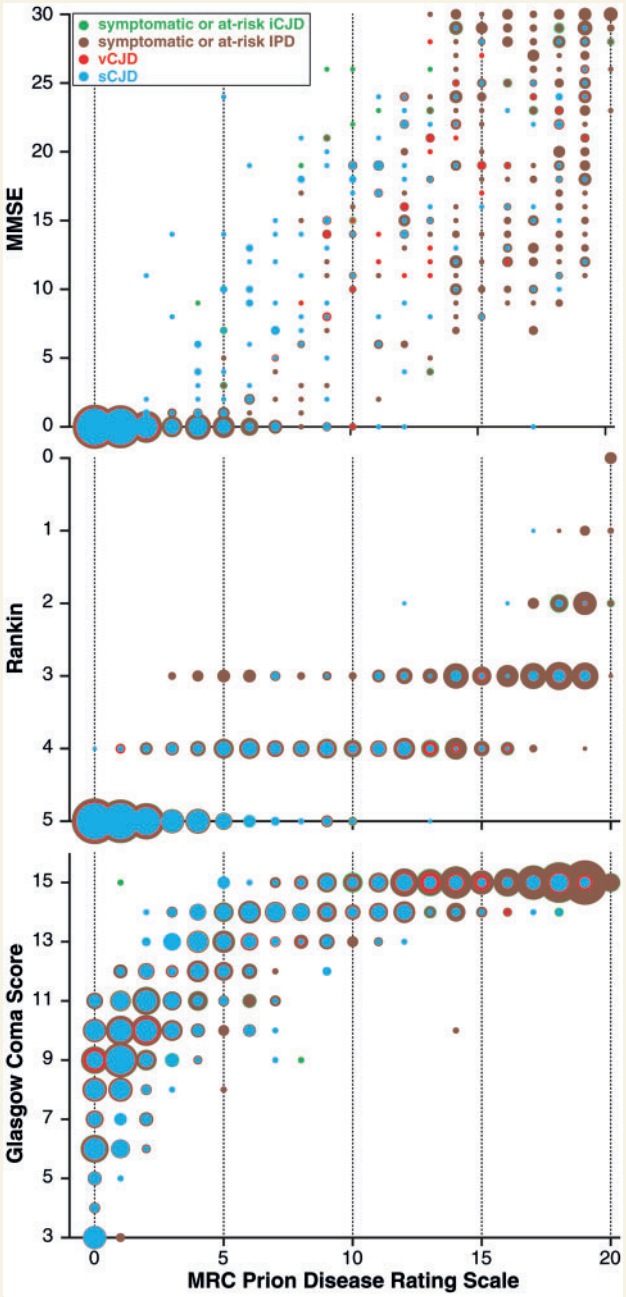

Figure 2 Correlation of the MRC Scale with other commonly used rating scales. The area coloured for each circle is proportional to the total number of patients from each aetiological group of prion disease with these scores. These plots illustrate the relative absence of floor effect with the MRC Scale when compared with the Rankin or the Mini-Mental State Examination (MMSE) scales, but not the Glasgow Coma Score, as patients with minimum/worst score on the Mini-Mental State Examination and Rankin can still be distinguished with the MRC Scale. In addition these illustrate the multidimensionality of the inherited prion disease (brown) group in that some patients decline in function with normal Mini-Mental State Examination (typically patients with Gerstmann-Straussler-Scheinker disease) and others decline in Mini-Mental State Examination with high levels of function (typically 6-octapeptide repeat insertion mutation patients). iCJD = iatrogenic CJD; IPD = inherited prion disease; sCJD = sporadic CJD; vCJD = variant CJD.

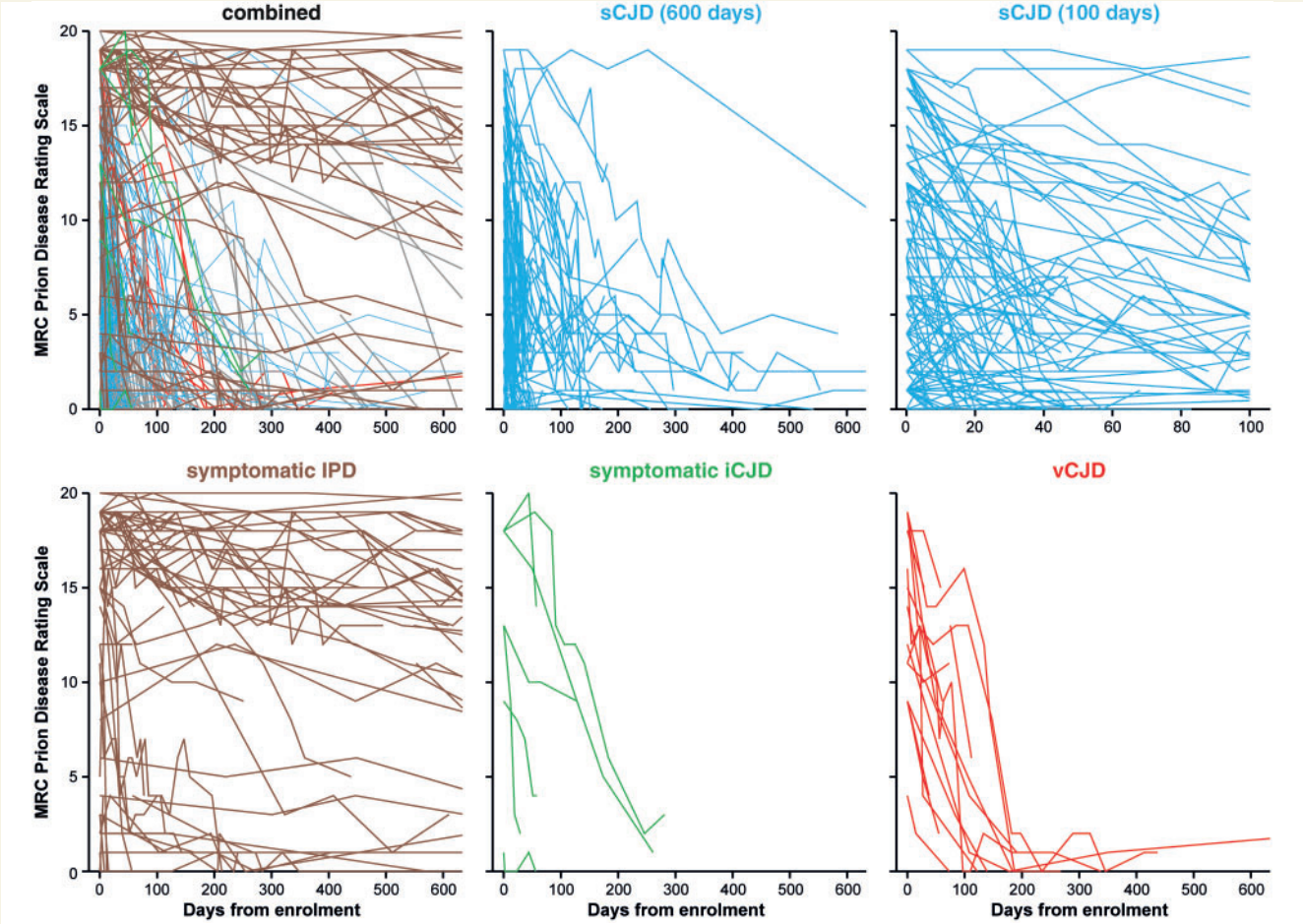

Figure 3 Trajectories of change in patients up to 600 days post-enrolment for all patients (combined), sporadic CJD (sCJD), inherited prion disease (IPD), iatrogenic CJD (iCJD) or variant CJD (vCJD) only. Three broad patterns are seen: a slow decline in inherited prion disease patients, a rapid and somewhat variable decline in all aetiological groups, and a pattern of decline followed by a preterminal plateau at low levels of function in all aetiological groups. All sporadic CJD trajectories are also shown for the first 100 days only for clarity of short duration cases. Data are also available up to 10 years in some patients with inherited prion disease.

Scale with Mini-Mental State Examination, Rankin and Glasgow Coma Score across all assessments at which these scales were acquired. This illustrates the novel scale's relative resistance to floor effects, as patients at minimum (or unrecordable) score for Mini-Mental State Examination or Rankin can still be separated by the novel scale. Whilst Glasgow Coma Score has the ability to separate patients scoring zero on the current scale (Glasgow Coma Score ranging 3–11), these distinctions are difficult to interpret clinically and would require a clinical examination.

The natural history of prion diseases

Figure 3 shows individual patient trajectories for the MRC Scale over time, colour-coded by disease type. Patients in Stratum 2 have been included for comparison purposes with the caveat that the MRC Scale is not measuring a single progression construct in these patients. These plots illustrate the remarkable heterogeneity of disease progression between, and to a lesser extent within, prion disease types. Several distinct patterns of progression are apparent: rapid decline over weeks or a few months (mostly patients with sporadic CJD) and slow decline over years (almost exclusively patients with inherited prion disease). A slightly less rapid decline is observed in patients with sporadic CJD who are heterozygous at *PRNP* codon 129, patients with variant CJD and iatrogenic CJD and some patients with inherited prion disease (Figs 3 and 4). In all groups, decline measured with the MRC Scale generally appears to be linear.

In most rapidly progressing patients, death occurs shortly after a very low score is reached, but in some there is an extended 'preterminal plateau' phase at a very low score. A prion disease clinical trial design needs to consider these different patterns of decline.

Figure 5 illustrates the typical progression of a patient with sporadic CJD through the functional/cognitive milestones of the MRC Scale, and the spread of difficulty of items in the five measured domains. These observations are in keeping with our clinical experience of these patient groups and are consistent with the validity of the MRC Scale.

Discussion

In this paper we have sought to overcome a fundamental obstacle on the route to developing effective treatments for prion disease: the lack of both a validated outcome measure for clinical trials and a large resource of clinical progression data. Using a range of complementary approaches, and taking into account analysis of rating scales data from the PRION-1 trial (Mead *et al.*, 2011), we have developed, refined and validated a bespoke rating scale in the context of the Cohort study, the largest prospective clinical study of the natural history of prion disease. Our outcome measure aims to maximize the likelihood of future trials giving a clear answer on therapeutic efficacy. This study also illustrates that the advanced neurodisability at referral of patients with sporadic CJD to our clinical research team is a major outstanding problem for UK clinical trials; improved early diagnosis and referral will be key to success of clinical trials.

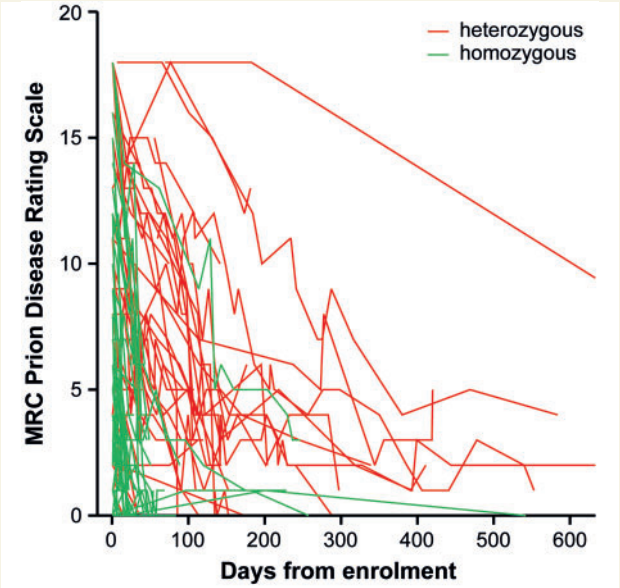

Figure 4 Trajectories of patients with sporadic CJD either homozygous (129MM or 129VV) at codon 129 of the prion protein gene or heterozygous (129MV). This genetic factor appears to be a strong determinant of rates of decline.

Prion disease trials have and will continue to use survival as a key outcome measure, but this has major limitations. It does not directly measure progression of disease, as patients may survive for long periods in a very advanced stage of disease or may die before, at or after reaching the end stages of disease progression (e.g. due to aspiration pneumonia). Existing rating scales, which are well validated in other neurological settings, are far less well suited to prion disease. For example, the Mini-Mental State Examination and Alzheimer's Disease Assessment Scale – cognitive subscale, routinely used as an outcome measure in Alzheimer's disease clinical trials, suffer from a marked floor effect (Mead *et al.*, 2011) and fail to capture the profound physical impairments that are fundamental features of these diseases and may be present despite preserved cognitive function. Cognitive decline is itself a fundamental feature of prion diseases, and we have included items assessing cognitive function in a way that is more robust in this population, by assessing carer-reported level of function.

The remarkable clinical heterogeneity of prion disease, combined with its rarity, represents a major challenge to trial design. Very inclusive enrolment criteria will maximize patient numbers,

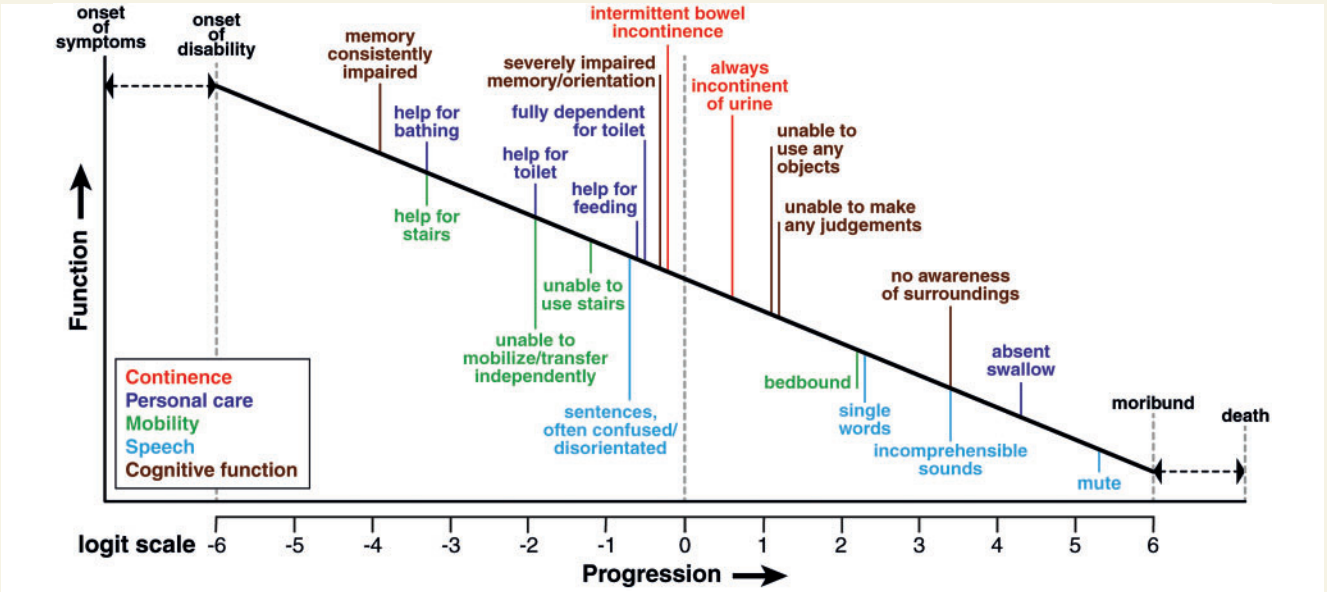

Figure 5 Schematic of the pattern of decline in a patient with sporadic CJD that would be most consistent with the Rasch model. Progression is represented by the logit scale used in the Rasch analysis reflecting the relative difficulties of the thresholds that comprise the MRC Scale. This diagram illustrates the validity of the progression construct, as the ordering of the items is consistent with clinical experience and there is a reasonable spread of item difficulty in different functional domains.

but may have a paradoxically negative effect on statistical power if they greatly increase variability, or reduce the possible benefits that can be observed either due to permanent neurological damage that cannot be rescued, or due to relatively preserved functioning at enrolment. Designing a single outcome measure that can capture rapid global decline in a patient with sporadic CJD and changes in patients with some slowly progressive forms of inherited prion disease has not been possible. Patients with inherited prion disease are highly variable in clinical presentation with predominantly mobility or predominantly cognitive progression. It is likely that the detection of subtle changes required to define the onset of disease in inherited prion disease will require analysis of neuropsychological testing that is ongoing in the Cohort study. Our findings suggest that, for maximum efficiency, the main group that should be targeted for future trial recruitment should be those Stratum 1 patients not already at very low levels of functioning. This group of patients were also most likely to choose to take the investigational drug quinacrine in PRION-1 (Collinge *et al.,* 2009).

While we have tried to design a scale that reflects the priorities of patients and their carers, a compromise must be struck with methodological concerns. Our scale does not include any direct assessment of neuropsychiatric symptomatology, sleep disturbance or movement disorder (e.g. myoclonus), all of which are common features of prion disease and of concern to patients and carers. These features have limited value as markers of disease progression however, as they often fluctuate through the course of the disease, may improve in the later stages, and may be significantly affected by non-disease modifying treatments (e.g. benzodiazepines, anticonvulsants). As such they are not included in our outcome measure, but we are currently investigating them using other methods (e.g. use of the Neuropsychiatric Inventory), and these data will be reported elsewhere. Analyses of the baseline predictors of rates, patterns of decline, simulations of clinical trials to estimate power, and fitting of linear mixed models are ongoing.

We have built up a large and detailed natural history data set, with the intent that this can act as a supplementary historical control group against which to compare treatment groups in future trials, with potentially large benefits for statistical power. To this end, the data set can be made available to other physicians conducting clinical trials, and we encourage research groups worldwide to take advantage of this in planning clinical trials. To enable this direct comparison of results from different studies, we propose that the MRC Prion Disease Rating Scale be adopted as a standard outcome measure for prion disease clinical research.

Clinical research into rare diseases is extremely challenging for logistic, statistical and financial reasons, but it is essential that we work towards overcoming these challenges. Considered together, 'rare diseases' make up a significant proportion of the burden of neurological disease, and it is essential to collect systematic data on which to base the treatment of these patients. Studying rare diseases can often provide valuable insights into more common conditions. In the wider field of neurodegeneration there is great interest in the hypothesis that templated protein misfolding mechanisms (referred to as 'prion-like') may be fundamental to a wide range of other conditions including Alzheimer's and Parkinson's diseases. Demonstrating a disease-modifying effect of a therapeutic agent in prion disease may therefore lead to insights with wider implications.

Acknowledgements

Prion-1 Trial steering committee—D Armstrong, I Chalmers, L Firkins, F Certo, J Collinge, J Darbyshire, C Kennard, A Kennedy (until 2004), J Ironside, A MacKay, H Millar, J Newsom-Davies (deceased), J Nicholl, J Stephenson, M Wiles, S Wroe. *Cohort steering committee*- C Kennard, P Mills, V Farewell, P Chinnery, A Mackay, E Riboli, R Knight. We thank all the individuals, their carers, and their families, who took part in the PRION-1 and Cohort studies and UK neurologists and the National CJD Surveillance Unit for referring patients. We thank the former Chief Medical Officer Sir Liam Donaldson, officials at the Department of Health, Medical Research Council Research Management Group staff, co-chairs of the PRION-1 and Cohort trial steering committee, and our colleagues at the National CJD Research Surveillance Unit for establishing the National CJD referral arrangements, without which these studies would not have been possible. We thank all past and present colleagues at the National Prion Clinic (formerly at St Mary's Hospital, London and now at the National Hospital for Neurology and Neurosurgery, Queen Square, London) in particular Durre Siddique, Suvankar Pal, Thomas Webb, Diego Kaski, Dillip Gallujipali, Kathryn Prout, Nora Heard, Clare Morris, Rita Wilkinson, Chris Rhymes, Suzanne Hampson, Claire Petersen, Chris Carswell, Joanna Field, Elisabeth Morgan, Ekaterina Kassidi, and Colm Treacy and at the Medical Research Council Clinical Trials Unit, in particular Michael Ranopa, Geraldine Keogh, Moira Spyer, Debbie Johnson, Liz Brodnicki, and Patrick Kelleher. We thank Professor A. Tennant and Dr E.M.A. Hensor for guidance on Rasch analysis. We acknowledge helpful discussion with Dr Diane Playford.

Ray Young assisted with figure design.

Funding

The Cohort study was funded by the Department of Health (England and Wales) with additional support from the Medical Research Council. Some of this work was undertaken at UCLH/UCL who received a proportion of funding from the Department of Health's NIHR Biomedical Research Centres funding scheme.

Conflict of interest

J.C. is a director and shareholder of D-Gen Ltd, an academic spin-out company working in the field of prion disease diagnosis, decontamination and therapeutics. No other author has a conflict of interest.

References

Andrich D, Sheridan B, Luo G. Perth: RUMM Laboratory; 2030. 2008. Rasch Unidimensional Measurement Model. 2008 [Computer software].

Collin C, Wade DT, Davies S, Horne V. The Barthel ADL Index: a reliability study. Int Disabil Stud 1988; 10: 61–3.

Collinge J. Prion diseases of humans and animals: their causes and molecular basis. Ann Rev Neurosci 2001; 24: 519–50.

Collinge J, Gorham M, Hudson F, Kennedy A, Keogh G, Pal S, et al. Safety and efficacy of quinacrine in human prion disease (PRION-1 study): a patient-preference trial. Lancet Neurol 2009; 8: 334–44.

Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975; 12: 189–98.

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; 6: 1094–105.

Klohn PC, Farmer M, Linehan JM, O'Malley C, de Marco MF, Taylor W, et al. PrP antibodies do not trigger mouse hippocampal neuron apoptosis. Science 2012; 335: 52.

Mahoney FI, Barthel D. Functional evaluation: the Barthel Index. Md State Med J 1965; 14: 56–61.

Mallucci G, Dickinson A, Linehan J, Klohn P, Brandner S, Collinge J. Depleting neuronal PrP in prion infection prevents disease and reverses spongiosis. Science 2003; 302: 871–4.

Mallucci G, Ratté S, Asante E, Linehan J, Gowland I, Jefferys JG, et al. Post-natal knockout of prion protein alters hippocampal CA1 properties, but does not result in neurodegeneration. EMBO J 2002; 21: 202–10.

Mallucci G, White MD, Farmer M, Dickinson A, Khatun H, Powell AD, et al. Targeting cellular prion protein reverses early cognitive deficits and neurophysiological dysfunction in prion-infected mice. Neuron 2007; 53: 325–35.

Mead S, Ranopa M, Gopalakrishnan GS, Thompson AG, Rudge P, Wroe S, et al. PRION-1 scales analysis supports use of functional outcome measures in prion disease. Neurology 2011; 77: 1674–83.

Mead S, Rudge P, Hudson F, Walker S, Darbyshire J, Collinge J. Clinical trials and methodological problems in prion diseases reply. Lancet Neurol 2009; 8: 782–3.

Morris JC. The Clinical Dementia Rating (CDR): current version and scoring rules. Neurology 1993; 43: 2412–14.

Nicoll AJ, Collinge J. Preventing prion pathogenicity by targeting the cellular prion protein. Infect Disord Drug Targets 2009; 9: 48–57.

Nicoll AJ, Trevitt CR, Tattum MH, Risse E, Quarterman E, Ibarra AA, et al. Pharmacological chaperone for the structured domain of human prion protein. Proc Natl Acad Sci USA 2010; 107: 17610–15.

Rankin J. Cerebral vascular accidents in patients over the age of 60. Scott Med J 1957; 2: 127–36.

Teasdale G, Jennett B. Assessment of coma and impaired consciousness - practical scale. Lancet 1974; 2: 81–4.

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; 57: 1358–62.

White AR, Enever P, Tayebi M, Mushens R, Linehan J, Brandner S, et al. Monoclonal antibodies inhibit prion replication and delay the development of prion disease. Nature 2003; 422: 80–3.

WHO. The revision of the surveillance case definition for variant creutzfeldt-jakob disease (vCJD). WHO manual for surveillance of human transmissible spongiform encephalopathies 2001.

Zerr I, Kallenberg K, Summers DM, Romero C, Taratuto A, Heinemann U, et al. Updated clinical diagnostic criteria for sporadic Creutzfeldt-Jakob disease. Brain 2009; 132: 2659–68.