Available online at www.sciencedirect.com

SciVerse ScienceDirect

Neuromuscular Disorders 23 (2013) 256–264

www.elsevier.com/locate/nmd

The Rasch-built Pompe-specific Activity (R-PAct) scale

N.A.M.E. van der Beek [a,b,*], M.L.C. Hagemans [b], A.T. van der Ploeg [b], P.A. van Doorn [a], I.S.J. Merkies [a,c,d]

[a] Department of Neurology, Center for Lysosomal and Metabolic Diseases, Erasmus MC University Medical Center, Rotterdam, The Netherlands
[b] Department of Pediatrics, Center for Lysosomal and Metabolic Diseases, Erasmus MC University Medical Center, Rotterdam, The Netherlands
[c] Department of Neurology, Spaarne Hospital, Haarlem, The Netherlands
[d] Department of Neurology, Maastricht University Medical Center, The Netherlands

Received 8 August 2012; accepted 30 October 2012

Abstract

We constructed a patient-based interval scale using Rasch analysis, specifically suited to quantify the effects of Pompe disease on patient's ability to carry out daily life activities and their social participation: Rasch-built Pompe-specific Activity scale. Between July 2005 and April 2011, 186 patients aged 16 or older, participated to develop this scale. External construct validity was determined through correlations with the MRC sumscore and Rotterdam Handicap Scale. Furthermore, test–retest reliability was determined in a subgroup of 44 patients. Finally, individual person-level responsiveness was used to determine the proportion of patients demonstrating significant improvement or deterioration during their natural disease course, or during treatment with enzyme replacement therapy. Of the original 49 items, 31 were removed after investigation of model fit, internal reliability, threshold examination, item bias, and local dependency. The remaining 18 items were ordered on a linearly weighted scale and demonstrated good discriminative ability (Person Separation Index 0.96), external construct validity (intraclass correlation coefficient (ICC) for MRC sumscore 0.82, and for the Rotterdam handicap scale 0.86), reliability of person's location (ability comparison: ICC 0.95), and responsiveness. We therefore conclude that the R-PAct scale enables us to accurately detect limitations in activities and social participation throughout the entire disease spectrum in patients with Pompe disease.

Keywords: Pompe disease; Acid α-glucosidase; Patient reported outcome measures; Disability measure, Rasch analysis

1. Introduction

Pompe disease – also named glycogen storage disease type II – is an inherited metabolic disorder in which partial or total absence of the enzyme acid α-glucosidase causes intra-lysosomal accumulation of glycogen in many tissues [1]. The clinical spectrum ranges from the rapidly progressive classic infantile phenotype, which – when not treated with enzyme replacement therapy – leads to death within the first year of life [2,3], to a more slowly progressive phenotype that primarily affects skeletal and respiratory muscles [4–6]. Progressive muscle weakness eventually leads to wheelchair and ventilator dependency in a substantial amount of patients.

As a consequence, Pompe disease strongly affects patients' ability to carry out daily life activities and influences their social participation. Quantifying these aspects is important for the management of individual patients and for evaluating effects of enzyme replacement therapy (ERT) [7,8] or future treatment modalities. At present, limitations in activities and social participation are often assessed by non-specific functional tests such as the 10-metre walk test and the six-minute walk test [7,8], or ordinal measurement scales such as the Rotterdam

* Corresponding author. Address: Department of Neurology, Center for Lysosomal and Metabolic Diseases, Erasmus MC University Medical Center, 's-Gravendijkwal 230, 3015 CE Rotterdam, The Netherlands. Tel.: +31 10 7033758; fax: +31 10 7035927.
E-mail address: n.beek@erasmusmc.nl (N.A.M.E. van der Beek).

0960-8966
http://dx.doi.org/10.1016/j.nmd.2012.10.024

handicap scale or the gross motor function measure [9–11]. It is now realised that these ordinal scales are prone to differential sensitivity – meaning that a one-point change in score at the centre of the scale may not be the same as a one-point change at the extremes [12,13]. Therefore, for health evaluation, a modern scientific approach transforming ordinal scores into a linearly weighted measure is required.

We thus developed a patient-based interval scale (Rasch-built Pompe-specific Activity scale: R-PAct scale) using Rasch analysis [14,15]. Subsequently, we evaluated its validity, reliability and responsiveness. It is to be expected that this measurement scale, based on patients' experiences of limitations in daily life, will have a high discriminatory capacity and will be able to measure changes in functional status, which aids in follow-up of the natural disease course or in the evaluation of therapeutic efficacy throughout the entire spectrum of disease severity.

2. Patients and methods

2.1. Study population and procedures

Between July 2005 and April 2011, 186 patients aged 16 or older, participated to develop the R-PAct scale. Patients were recruited though patient organisations affiliated with the International Pompe Association (IPA) in Canada, the Netherlands, the United Kingdom, and the United States [6]. In addition, patients were recruited through neuromuscular centres in the Netherlands and Belgium. Forty–four of the Dutch patients completed the scale a second time 2–4 weeks later for test–retest reliability studies. To evaluate responsiveness of the R-PAct scale, patients were followed longitudinally up to 36 months during the natural disease course or during treatment with enzyme replacement therapy. Patients from the international patient cohort completed the scale every year, while the Dutch and Belgian patients completed the scale every 3–6 months. The studies were approved by the Medical Ethics Committee of Erasmus MC University Medical Center. All patients gave written informed consent.

2.2. Scale development

2.2.1. Preliminary R-PAct questionnaire

A preliminary R-PAct questionnaire was developed taking the following steps:

(1) To cover the widest range of physical functioning, activities, and participation skills important for patients with Pompe disease, data from an international patient survey among 263 patients with Pompe disease were used as the basis for construction of the current questionnaire [6]. In this survey, patients provided information about their disease history and current status by means of self-reported questionnaires. Firstly, an inventory was made of the answers on the following open question from this patient survey: "what are your most important (limiting, annoying) problems?" Secondly, answers to the question "can you describe your walking problems?" were used to refine the items on walking. Thirdly, patients were asked to indicate other items that were important for their daily life but that were not discussed in the questionnaires.
(2) Additional items based on the Rotterdam Handicap Scale (1 item) [10] and paediatric evaluation of disability inventory (8 items) [16], thought to be of importance by expert judgement, were included.
(3) This preliminary list comprising 136 items was classified according to the International classification of functioning, disability and health (ICF) into the domains impairment, activities, and participation [17].
(4) A panel of experts consisting of senior staff members from our departments of neurology, paediatrics, internal medicine, and clinical genetics, all involved in research projects and treatment of Pompe disease, discussed the items and their classification. Changes were implemented according to their suggestions: items that were almost identical were merged, and items that were mentioned only occasionally were left out. After reaching consensus, a list of 49 activity and participation items was selected for the purpose of the current study. All items had five response options: (0) unable to perform; (1) able to perform, but with great difficulty; (2) able to perform, but with some difficulty; (3) able to perform, but with little difficulty; (4) easy to perform, without difficulty; or 'not applicable'.
(5) The questionnaire was tested in a group of ten healthy subjects. Based on their comments, changes were made to prevent overlap and to improve clarity.
(6) For use in English speaking patient groups, this final questionnaire was translated and back-translated by two independent certified translators according to published guidelines [18].

The items of the preliminary R-PAct questionnaire are listed in Webappendix 1.

2.3. Final R-PAct scale

2.3.1. Rasch analysis

Since a sample size of approximately 250 is needed to adequately estimate item difficulty [26], we decided to stack the data of the first ($n = 186$) and second ($n = 44$) assessments, controlling for 'time factor' as possible confounding factor [27], leading to a total number of 230 records to be examined. In the model construction, items scored as 'not applicable' were interpreted as missing data. Items with more than 10% missing values and questionnaires with more than 10% unanswered items were omitted as a quality control procedure. Thereafter, the remaining response

values of the preliminary R-PAct were analysed using Rasch unidimensional measurement models (RUMM 2030) [19]. Through Rasch analysis, ordinal scores are transformed into interval measures, placing both item and person parameter estimates on the same log-odds units (logit) scale, which allows for linear transformation of the raw scores [14,15]. A detailed description of the statistical modelling procedures has been provided elsewhere, also specifically for neurologists [20–22]. Briefly, the following Rasch model requirements were checked:

(A) *Fit statistics and fit residuals:* To test whether the data meet the model expectations, three overall fit statistics were considered. Two are item–person interaction statistics, expressed as *z*-scores: if the items and persons fit the model, a mean around zero and a standard deviation of 1 would be expected. The third is an item–trait interaction statistic, reported as chi-square (χ^2): a non-significant chi-square reflects the required property of invariance. Additionally, individual person-fit statistics and item-fit statistics were examined as residuals, and by using a chi-squared statistic. Residuals between ± 2.5 are considered adequate fit to the model, whilst a significant χ^2 points to misfit.

(B) *Internal reliability:* This was measured by the Person Separation Index (PSI). A value of $\geqslant 0.7$ indicates that the scale is able to differentiate at least two groups of patients, and is generally considered to be acceptable [23].

(C) *Threshold examination:* The point between two adjacent response categories where both responses are equally probable is called the "threshold". Difficulty to discriminate between response options – for example due to too many possibilities – may lead to disordered thresholds. The overall model fit may improve by merging of categories.

(D) *Item bias:* This was assessed by means of differential item functioning (DIF) [24]. DIF occurs when the probability of responding to an item is systematically different between groups with equal levels of disability but differences in another characteristic (e.g. age). This was examined by analysis-of-variance for the following arbitrarily chosen personal factors, aiming for an equal distribution of patients among the categories: (1) age (<40 years, 40–50 years, 50–60 years, or $\geqslant$60 years); (2) gender (male or female); (3) duration of symptoms (<5 years, 5–10 years, 10–20 years, or $\geqslant$20 years); (4) country of assessment (Belgium, Canada, the Netherlands, UK, or USA); and (5) language (Dutch or English).

(E) *Local dependency:* Local dependency arises when items are linked. For example, a patient who is unable to walk 100 m, will also be unable to walk 1 km. An inter-item residual correlations $\geqslant 0.3$ indicates local dependency [25].

(F) *Unidimensionality:* This was tested by a principal component analysis of the residuals, needed to support the assumption of local independence and, consequently, the unidimensionality of the scale [21].

Throughout the analyses, we continuously monitored whether Rasch model criteria were met: items that did not fulfill these requirements were removed, or adjusted to fit the model, while monitoring changes and fit statistics of the individual remaining items and the overall model fit. As a last step, the calculated Rasch person locations (in logits) were transformed into a more understandable centile metric ranging from 0 (most severe activity and participation restrictions) to 100 (no activity limitations and participation restrictions).

2.3.2. Validity

The external construct validity was assessed by correlations between the final R-PAct scale and the Medical Research Council (MRC) [28] sumscore and Rotterdam Handicap Scale (RHS) [10]. To obtain the most appropriate graphical model fit, regression analysis with restricted cubic spline functions was performed on the summed raw R-PAct scores: the intraclass correlation coefficient (ICC) is reported [29].

2.3.3. Medical Research Council (MRC) score

Through manual muscle testing using the MRC grading system [28], we assessed the degree of skeletal muscle weakness in the 88 patients participating in the ongoing study on the natural course of Pompe disease in the Netherlands. Since, in patients with Pompe disease, muscle weakness is present predominantly in the proximal muscle groups ('limb-girdle' weakness), leaving the distal muscle groups relatively unaffected until the late stages of the disease, a sumscore was calculated for the following muscles or muscle groups: neck extensors, neck flexors, and bilateral shoulder abductors, elbow flexors, elbow extensors, hip flexors, hip abductors, knee flexors and knee extensors (range 0 (paralytic)–80 (normal strength)).

2.3.4. Rotterdam Handicap Scale (RHS)

The RHS was developed originally for measuring handicap in patients with immune-mediated polyneuropathies [10], and was proven to be useful for examining limitations in activities and participation in patients with Pompe disease [9]. The scale consists of nine questions on the topics of mobility indoors and outdoors, kitchen tasks, domestic tasks indoors and outdoors, leisure activities indoors and outdoors, travelling, and work/study. The total score ranges from 9 to 36, with higher values representing a lower level of handicap.

2.3.5. Reliability

To investigate whether the hierarchy of patients' ability location was consistent over time, test–retest reliability

studies were performed [30]. Reliability was quantified by calculating the intraclass correlation coefficient using a one-way random effects analysis-of-variance (ANOVA) model for group comparison.

2.3.6. Responsiveness

Responsiveness was calculated at the individual person-level, since modern clinimetric methods have demonstrated that the standard error (SE) around an individual patients' ability level, and therewith the clinical importance of changes within a patient over time, may vary across the range of an outcome measure [31,32]. As a measure for individual responsiveness, the minimal clinically important difference-standard error (MCID-SE: individual change/standard error of difference (SE_{diff})) was calculated for every participating patient at each assessment [33]. To obtain individual SEs, all data were subjected to the RUMM 2030 model first, hereby creating the location of each patient (in logits) with the corresponding SE (also in logits). The cut-off value for a clinically important change – improvement or deterioration – was defined at $\pm 1.96 \times SE$. Subsequently, Kaplan–Meier curves were applied to estimate the cumulative proportion of patients demonstrating significant improvement or significant deterioration over time (3, 6, 9, 12,... up to 36 months), stratified for 'treatment' (natural disease course against enzyme replacement therapy). The log-rank test was used to examine possible group differences.

2.4. Statistics and software

Rasch analyses were performed with the partial credit model as default (RUMM2030) [19]. Further analyses were undertaken using Stata Statistical Software for Windows XP (version 11.0, StataCorp, Texas, USA). Throughout the analyses, Bonferroni corrections were applied to adjust the *p*-values for multiple testing [34].

3. Results

3.1. Study population

The study population comprised patients from the Netherlands ($n = 94$), the USA ($n = 65$), the UK ($n = 18$), Canada ($n = 6$), and Belgium ($n = 3$). Fifty-one percent was female. The median age at which patients had been diagnosed was 37 years (range 1–67 years). Median age at first investigation was 50 years (range 23–85 years), and median disease duration 11 years (range 0–33 years). Thirty-seven percent of patients were fully ambulatory, 16% used walking devices, and 47% of all patients was either partially or permanently wheelchair dependent. Forty-five percent of patients used mechanical ventilatory support. Patients from the United States, United Kingdom or Canada had a longer disease duration ($p < 0.01$) and were more frequently ventilator dependent than the Dutch or Belgian patients ($p < 0.01$).

3.2. Rasch analysis

3.2.1. Initial analysis on the preliminary R-PAct questionnaire

The preliminary R-PAct questionnaire (see Appendix 1 for full list of items) did not meet all Rasch model expectations (Table 1, initial analysis).

3.2.2. Data handling to fit the Rasch model

(1) Thirty items of the preliminary R-PAct questionnaire demonstrated disordered thresholds, particularly in the mid-response area (response options 1–3, Fig. 1). The remaining 19 items had thresholds very adjacent to each other in the mid-response area. Based on these observations, we decided to rescore all items into three response categories as follows: (0) = (0) unable to perform; (1–3) = (1) able to perform, but with difficulty; and (4) = (2) able to perform without difficulty.
(2) Inspection of individual item-fit and individual person-fit statistics showed that 14 items demonstrated misfit to the model, 13 having a significant chi-square probability, and one having fit residuals exceeding −2.5. These items were removed one by one, continuously checking the class intervals, statistical control panel and possible changes on other Rasch requirements.
(3) Thirteen items showed item-bias: six items with regard to the personal factor 'country', three for 'language', two for 'age', and two for 'duration of symptoms'. These items were removed from the analyses.
(4) To identify possible local dependency, all pairs of items with correlations above 0.28 were evaluated, starting with the highest correlations. Of each item-pair, the item showing the least clinical relevance

Table 1
Summary statistics of Rasch analysis during construction of the R-PAct scale for Pompe patients.

Analysis	Item fit residuals		Person fit residuals		Item-trait χ^2 interaction		PSI	Unidimensionality *t*-tests (95% CI)
	Mean	SD	Mean	SD	DF	*p*-Value		
Initial[a] (preliminary R-PAct)	−0.254	1.780	−0.146	0.982	147	<0.0001	0.98	0.22 (0.187–0.243)
Final (R-PAct)	−0.353	0.826	−0.318	0.839	54	0.82	0.96	0.061 (0.032–0.090)

Abbreviations: χ^2, Chi square; DF, degrees of freedom; SD, standard deviation; CI, confidence interval; PSI, Person Separation Index.
[a] Thirty-one items were removed from the original preliminary R-PAct questionnaire. The remaining 18 items as part of the final R-PAct scale fulfilled all Rasch model expectations.

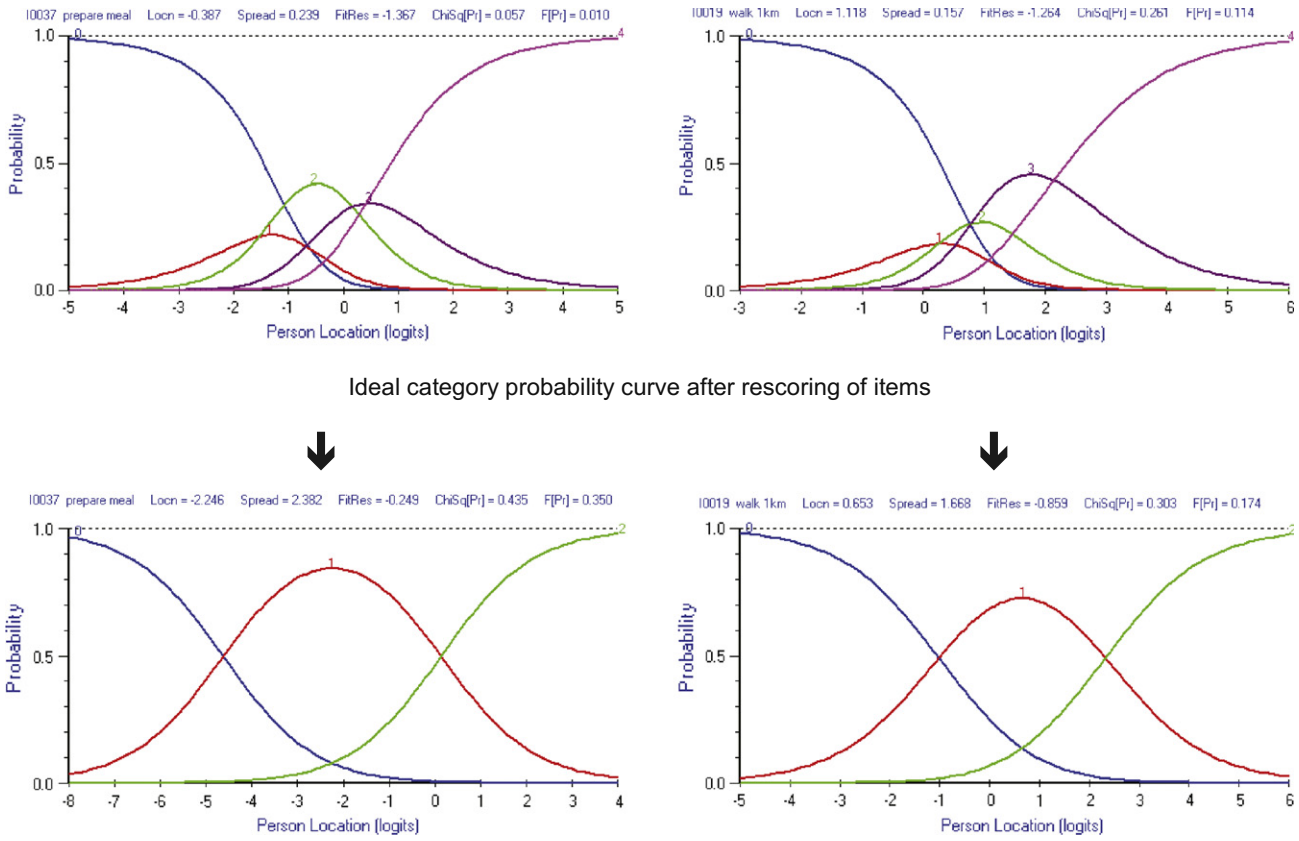
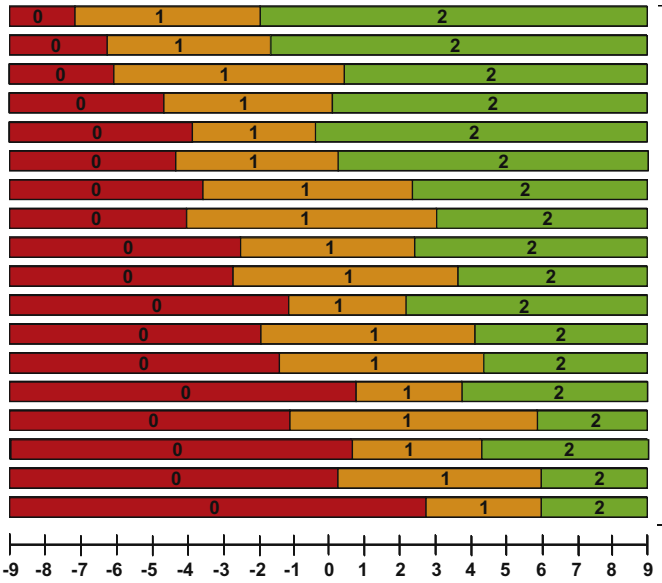


Fig. 1. Upper panels: two items (left: prepare a meal, right: walk 1 km) are presented as examples of reversed thresholds. Response categories, particularly in the mid response area (from 1 to 3) were not equally probable, indicating the inability of the patients to discriminate between these response options. Thirty of the initial items demonstrated this pattern. Lower panels: ideal probability curves after rescoring the items from 5 to 3 response options.



	Are you able to	Location	SE
1	comb your hair?	-4.533	0.186
2	eat?	-3.943	0.175
3	put on your trousers?	-2.878	0.172
4	prepare a meal?	-2.246	0.166
5	take a shower?	-2.131	0.155
6	grab an object above the head	-1.963	0.166
7	negotiate obstacles when walking?	-0.568	0.180
8	turn around in bed?	-0.453	0.200
9	walk on uneven ground?	0.001	0.167
10	stand up from a sitting position	0.536	0.185
11	walk 1 km outside?	0.653	0.153
12	walk 1 flight of stairs?	1.165	0.178
13	bend over and pick up object from the floor?	1.506	0.173
14	walk at a rapid speed	2.275	0.171
15	perform garden tasks?	2.415	0.188
16	practice a sport?	2.465	0.203
17	squat down and up?	3.164	0.181
18	run?	4.443	0.254

Fig. 2. Threshold distribution map and fit statistics for the 18-item final R-PAct scale, fulfilling all Rasch model expectations. Items are ordered by increasing difficulty. The easiest item turned out to be 'are you able to comb hair', the most difficult one 'are you able to run'. Red sections (0) = unable to perform, orange sections (1) = able to perform, but with difficulty, green sections (2) = able to perform, without difficulty.

Table 2
Nomogram allowing translation of summed raw scores of the final 18-items R-PAct scale (range 0–36) to an estimate of the Rasch person location (in logits) and a convenient centile metric (range 0–100). The corresponding logits in relation to the summed raw scores are provided by the RUMM software.

R-PAct summed raw score[a]	Rasch person location (logit)	Centile metric
0	−8.33	0
1	−7.25	7
2	−6.37	12
3	−5.66	17
4	−5.05	20
5	−4.53	24
6	−4.07	26
7	−3.65	29
8	−3.25	32
9	−2.88	34
10	−2.52	36
11	−2.18	38
12	−1.85	40
13	−1.54	42
14	−1.23	44
15	−0.92	46
16	−0.61	48
17	−0.31	50
18	0.00	52
19	0.31	54
20	0.63	56
21	0.95	58
22	1.29	60
23	1.63	62
24	1.99	64
25	2.34	66
26	2.69	68
27	3.04	70
28	3.40	73
29	3.76	75
30	4.14	77
31	4.55	80
32	5.00	83
33	5.50	86
34	6.09	89
35	6.84	94
36	7.79	100

[a] The nomogram can only be used when all questions have been completed by the patient.

(i.e. face validity and content validity), and the most over-discrimination or under-discrimination on its category probability curve was removed. In this way, four items were removed.

After completing these procedures, the final R-PAct scale, comprising 18 items, met all Rasch model expectations (Table 1, final analysis). In the final R-PAct scale, the item 'are you able to comb your hair?' was the easiest to perform whereas the item 'are you able to run?' turned out to be the most difficult task. Item difficulty ranged from −4.53 to 4.44 logits and patient ability level from −8.33 to 7.79 logits. Fig. 2 shows the threshold distribution map and the fit statistics for the 18 items of the final R-PAct scale.

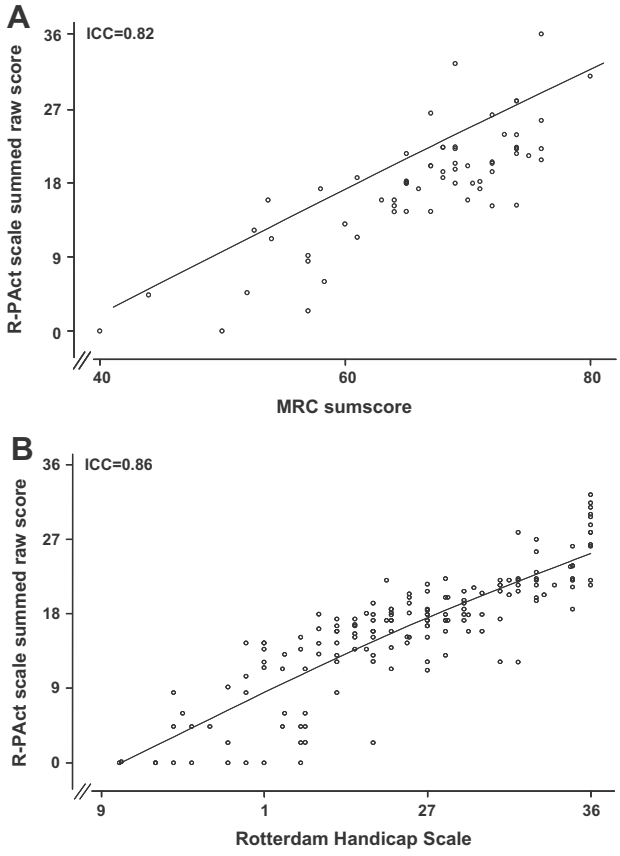


Fig. 3. Correlation between the R-PAct scale summed raw scores and MRC sumscore (A) and Rotterdam Handicap Scale (RHS) (B). ICC: intraclass correlation coefficient.

Table 2 provides a nomogram allowing the translation of the R-PAct summed raw scores to the calculated Rasch person location (in logits), and to an understandable centile metric. Fourteen patients (6%) could not perform any task at all (floor effect), and 1 patient (0.4%) was able to perform all activities without any difficulty (ceiling effect).

3.2.3. External construct validity

The R-PAct scale demonstrated strong correlations with the MRC sumscore (ICC 0.82) and RHS (ICC 0.86), reflecting good construct validity (Fig. 3). The discriminatory capacity of the R-PAct scale is clearly illustrated in Fig. 4. Patients with more disability, measured by the use of walking devices or wheelchair use, scored significantly lower on the R-PAct scale. The same pattern was seen when comparing patients using mechanical ventilation against those with no need for ventilation.

3.2.4. Reliability

The Person Separation Index was 0.96, demonstrating good internal consistency reliability. The test–retest reliability for person location was good as well: patient locations at the first and second assessment were nearly always within the 95% confidence intervals, reflecting ideal invariance (intraclass correlation coefficient 0.95) (Fig. 5).

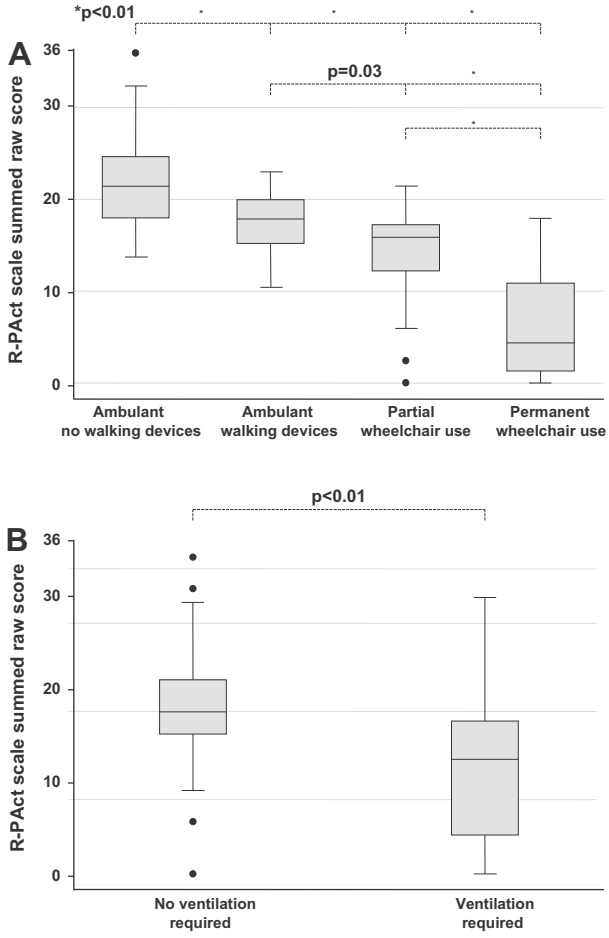


Fig. 4. Correlation between the R-PAct scale summed raw scores and mobility (A) and use of ventilatory support (B) illustrating the discriminatory capacity of the R-PAct scale. ICC: intraclass correlation coefficient.

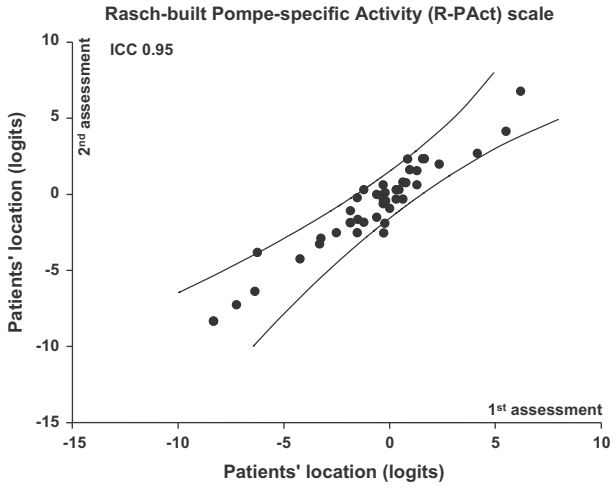


Fig. 5. Test–retest reliability studies between patients' ability location measured at the first and second assessment and the 95% confidence intervals (solid lines) of the ideal invariance. ICC: intraclass correlation coefficient.

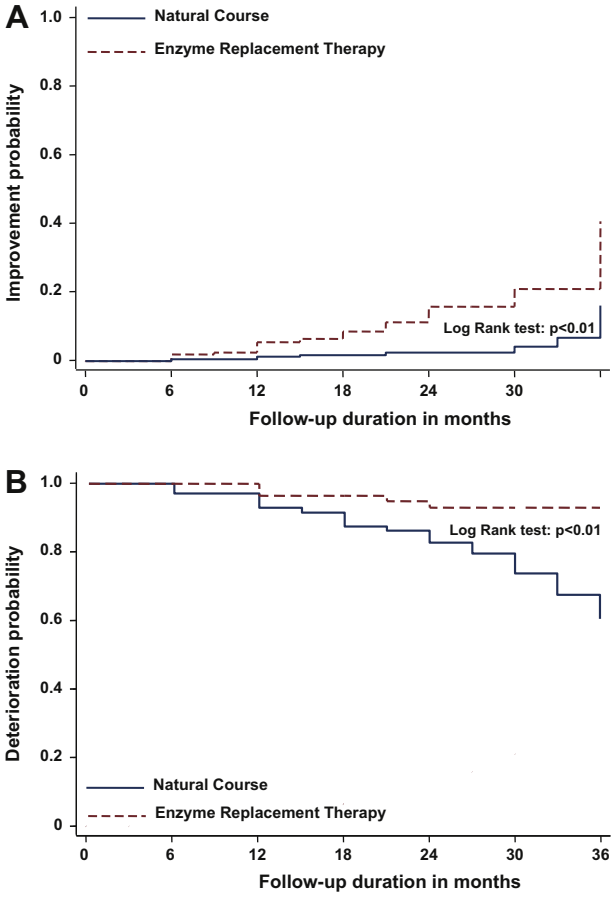


Fig. 6. Kaplan–Meier curves illustrating the proportion of Pompe patients demonstrating clinically meaningful improvement (A) or deterioration (B) during the natural disease course or during treatment with enzyme replacement therapy, applying the concept of minimal clinically important change (MCID) using a cut-off value of $\pm 1.96 \times$ SE.

3.2.5. Responsiveness

Longitudinal data on the natural disease course were available for 101 patients, while for 111 patients follow-up data following treatment with enzyme replacement therapy were evaluated. At 36 months, 21% of patients who were being treated with enzyme replacement therapy demonstrated a clinically important improvement at the *individual person-level*, against 7% of patients who were not being treated (i.e. natural disease course) ($p < 0.01$). By comparison, 33% of patients who did not receive enzyme replacement therapy demonstrated a clinically important deterioration compared to entry, while 7% patients who were being treated with enzyme therapy showed a clinically meaningful deterioration ($p < 0.01$). Fig. 6 shows the corresponding Kaplan–Meier curves.

4. Discussion

We developed and validated a new, self-reported, questionnaire designed specifically for use in patients with Pompe disease, based upon experiences from patients

about their most important and limiting aspects in daily life. The final 18-items R-PAct scale was able to measure across the spectrum of very severely to mildly affected patients without any relevant floor or ceiling effects, and showed good test–retest reliability. Furthermore, the final scale showed no item bias or local dependency and demonstrated acceptable unidimensionality. The high Person Separation Index, which indicates a strong ability of the scale to differentiate between patients with various degrees of ability, is proof of good internal consistency reliability. External construct validity was demonstrated by good correlations with the MRC sumscore and RHS, indicating that the R-PAct scale is capable of indirectly capturing physical impairments leading to problems in daily and social functioning [17]. Similar patterns between impairment, disability and handicap have been reported in patients with chronic immune-mediated neuropathies [35] and myotonic dystrophy [20].

This study contributes to the movement from classical test theory to modern test theory in the creation and evaluation of outcome measures in chronic neurological conditions. Incorporating MCID techniques will help in better understanding of the clinical relevance of changes in scores (e.g. defining responders *against* non-responders), rather than concentrating on statistical significance alone [38,39]. Traditional responsiveness indicators do not always provide information on the magnitude and direction of change (improvement, stable situation, or deterioration) for each individual patient. The Rasch method makes it possible to define 'response' at the individual person-level, taking into account individual standard errors changing according to patient's ability level, and may therefore have major implications also for future trials. Taking a cut-off value for MCID-SE of ± 1.96, the R-Pact scale was shown to detect clinically meaningful changes over time. This will be valuable in estimating the rate of disease progression, determining the best moment to initiate treatment, and evaluating therapeutic efficacy.

In recent years, several Rasch built measures of limitations in activities and participation have been constructed for patients with neuromuscular disorders (ACTIVLIM) [36], myotonic dystrophy type I (DM1-Activ) [20], immune-mediated neuropathies (R-ODS) [22], and also for children and adolescents with Pompe disease (Pompe-PEDI) [37]. Whereas the ACTIVLIM is a generic activity measure for patients with neuromuscular disorders, the DM1-Activ, R-ODS and Pompe-PEDI are disease-specific measurement instruments. The rationale for development of yet another scale was that we expected that a measurement instrument based on patients' own experiences would be the most appropriate to the patient group under study. It can be argued that for different patient groups selection of items would be different, or that the difficulty of the selected items is different. For example, items that are relevant for patients with a proximal myopathy do not necessarily apply to patients with pronounced distal skeletal muscle weakness. In part the current measurement scale overlaps with the existing measurement instruments: eight items of the final R-PAct scale are the same as in the DM1-Activ, and five items are shared with the ACTIVLIM. However, the estimated item difficulty differs substantially between the patient groups examined. Therefore, a disease-specific scale is more suited to estimating the impact of Pompe disease on daily life. In contrast to the Pompe-PEDI, which was developed specifically for children and adolescents with Pompe disease and especially takes the domains of mobility and self-care into account, the R-PAct scale is designed for use in patients of 16 years or older and also addresses aspects of social participation. We believe that the newly constructed scale gives more insight in the disabling impact of this disorder. Nevertheless, for a comprehensive overview of patients' functionality it should be used complementary to clinical evaluation methods measuring impairment, such as manual muscle testing, quantitative muscle testing, muscle function tests or pulmonary function testing.

Some limitations of the study need to be addressed. Ideally, a sample size of approximately 250 patients is needed to provide accurate model stability. This could only be reached by stacking of the data. Secondly, we used the MRC sumscore and RHS for establishing the external construct validity of the newly constructed scale, which are in fact ordinal summed scores. Recently, a revised MRC scoring system was developed, which is considered a substantial improvement in evaluating muscle strength [40]. It should now be determined whether this modified MRC scoring system is more appropriate for use in patients with Pompe disease.

In conclusion, the R-PAct scale enables us to accurately measure limitations in activities and restrictions in social participation throughout the whole spectrum of disease severity in patients with Pompe disease older than 16 years, and is able to capture clinically important changes over time. We therefore expect the R-PAct to be useful in future studies evaluating the natural disease course or therapeutic efficacy.

Acknowledgements

This study was a joint initiative of the International Pompe Association (IPA) and the Center for Lysosomal and Metabolic Diseases at Erasmus MC University Medical Center. We thank all patients for participating in the study, and Ria Broekgaarden, Marylyn House, and Allan Muir for their contribution to the organisational part of the study. We also thank Sonja van Nes for her critical reading of the manuscript.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.nmd.2012.10.024.

References

[1] van der Ploeg AT, Reuser AJ. Pompe's disease. Lancet 2008;372: 1342–53.

[2] Kishnani PS, Hwu WL, Mandel H, Nicolino M, Yong F, Corzo D. A retrospective, multinational, multicenter study on the natural history of infantile-onset Pompe disease. J Pediatr 2006;148:671–6.

[3] van den Hout HM, Hop W, van Diggelen OP, et al. The natural course of infantile Pompe's disease: 20 original cases compared with 133 cases from the literature. Pediatrics 2003;112:332–40.

[4] Engel AG, Gomez MR, Seybold ME, Lambert EH. The spectrum and diagnosis of acid maltase deficiency. Neurology 1973;23:95–106.

[5] Winkel LP, Hagemans ML, van Doorn PA, et al. The natural course of non-classic Pompe's disease; a review of 225 published cases. J Neurol 2005;252:875–84.

[6] Hagemans ML, Winkel LP, Hop WC, Reuser AJ, van Doorn PA, van der Ploeg AT. Disease severity in children and adults with Pompe disease related to age and disease duration. Neurology 2005;64: 2139–41.

[7] Strothotte S, Strigl-Pill N, Grunert B, et al. Enzyme replacement therapy with alglucosidase alfa in 44 patients with late-onset glycogen storage disease type 2: 12-month results of an observational clinical trial. J Neurol 2010;257:91–7.

[8] van der Ploeg AT, Clemens PR, Corzo D, et al. A randomized study of alglucosidase alfa in late-onset Pompe's disease. N Engl J Med 2010;362:1396–406.

[9] Hagemans ML, Laforet P, Hop WJ, et al. Impact of late-onset Pompe disease on participation in daily life activities: evaluation of the Rotterdam Handicap Scale. Neuromuscul Disord 2007;17:537–43.

[10] Merkies IS, Schmitz PI, van der Meche FG, Samijn JP, van Doorn PA. Psychometric evaluation of a new Handicap Scale in immune-mediated polyneuropathies. Muscle Nerve 2002;25:370–7.

[11] Russell DJ, Rosenbaum PL, Cadman DT, Gowland C, Hardy S, Jarvis S. The gross motor function measure: a means to evaluate the effects of physical therapy. Dev Med Child Neurol 1989;31:341–52.

[12] DeVellis RF. Classical test theory. Med Care 2006;44:S50–9.

[13] Tesio L. Functional assessment in rehabilitative medicine: principles and methods. Eura Medicophys 2007;43:515–23.

[14] Rasch G. Probabalistic models for some intelligence and attainment tests. Chicago: MESA Press; 1993.

[15] Bond TG, Fox CM. Applying the Rasch model: fundamental measurement in the human sciences. Mahwah: Lawrence Erlbaum Associates; 2001.

[16] Haley SM, Coster WJ, Ludlow LH, et al. Pediatric evaluation of disability inventory (PEDI). Development, standardization and administration manual. Boston (MA): New England Medical Center; 1992.

[17] World Health Organization. International classification of functioning, disability and health. Geneva: WHO; 2001.

[18] Streiner DL, Norman GR. Health measurement scales. A practical guide to their development and use. New York: Oxford University Press; 1998.

[19] Andrich D, Sheridan B, Luo G. Rasch models for measurement: RUMM2030. Perth: RUMM Laboratory; 2010.

[20] Hermans MC, Faber CG, de Baets MH, de Die-Smulders CE, Merkies IS. Rasch-built myotonic dystrophy type 1 activity and participation scale (DM1-Activ). Neuromuscul Disord 2010;20: 310–8.

[21] 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.

[22] van Nes SI, Vanhoutte EK, van Doorn PA, et al. Rasch-built overall disability scale (R-ODS) for immune-mediated peripheral neuropathies. Neurology 2011;76:337–45.

[23] Fischer WP. Reliability statistics. Rasch Meas Trans 1992;6:238.

[24] Dorans NJ, Holland PW. DIF detection and description: mantel-Haenszel and standardisation. In: Holland PW, Wainer H, editors. Differential item functioning. Hillsdale: Lawrence Erlbaum Associates; 1993. p. 33–6.

[25] Embretson SE, Reise SP. Item response theory for psychologists. New Yersey: Lawrence Erlbaum Associates; 2000.

[26] Linacre JM. Sample size and item calibration stability. Rasch Meas Trans 1994;7:28.

[27] Wright BD. Rack and stack: time 1 vs. time 2. Rasch Meas Trans 2003;17:905–6.

[28] Medical Research Council. Aids to examination of the peripheral nervous system. In: Memorandum no. 45. 1st ed. London: Her Majesty's Stationary Office; 1976.

[29] Herndon JE, Harrell FE. The restricted cubic spline hazard model. Commun Stat Theory Methods 1990;19:639–63.

[30] Wright BD, Stone MH. Best test design: Rasch measurement. Chicago: Mesa Press; 1979.

[31] Hobart JC, Cano SJ, Thompson AJ. Effect sizes can be misleading: is it time to change the way we measure change? J Neurol Neurosurg Psychiatr 2010;81:1044–8.

[32] Lai JS, Eton DT. Clinically meaningful gaps. Rasch Meas Trans 2002;15:850.

[33] Hobart J, Cano S. Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. Health Technol Assess 2009;13:1–200.

[34] Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. BMJ 1995;310:170.

[35] Merkies IS, Schmitz PI, van der Meche FG, Samijn JP, van Doorn PA. Connecting impairment, disability, and handicap in immune mediated polyneuropathies. J Neurol Neurosurg Psychiatr 2003;74: 99–104.

[36] Vandervelde L, van den Bergh PY, Goemans N, Thonnard JL. ACTIVLIM: a Rasch-built measure of activity limitations in children and adults with neuromuscular disorders. Neuromuscul Disord 2007; 17:459–69.

[37] Haley SM, Fragala MA, Aseltine R, Ni P, Skrinar AM. Development of a disease-specific disability instrument for Pompe disease. Pediatr Rehabil 2003;6:77–84.

[38] Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Control Clin Trials 1989;10:407–15.

[39] Merkies IS, van Nes SI, Hanna K, Hughes RA, Deng C. Confirming the efficacy of intravenous immunoglobulin in CIDP through minimum clinically important differences: shifting from statistical significance to clinical relevance. J Neurol Neurosurg Psychiatr 2010;81:1194–9.

[40] Vanhoutte EK, Faber CG, van Nes SI, et al. Modifying the Medical Research Council grading system through Rasch analyses. Brain 2012;135:1639–49.