1

2

3 **A functional scale for spinal and bulbar muscular atrophy**

4 **: cross-sectional and longitudinal study**

5

6 Atsushi Hashizume¹, Masahisa Katsuno¹, Keisuke Suzuki¹, Haruhiko Banno^{1, 2},

7 Noriaki Suga¹, Tomoo Mano¹, Amane Araki¹, Yasuhiro Hijikata¹, Christopher

8 Grunseich³, Angela Kokkinis³, Akihiro Hirakawa⁴, Hirohisa Watanabe¹, Masahiko

9 Yamamoto⁵, Kenneth H. Fischbeck³, and Gen Sobue¹

10

11 1. Department of Neurology, Nagoya University Graduate School of Medicine,

12 Nagoya, 466-8550 Japan

13 2. Institute for Advanced Research, Nagoya University, Nagoya, 464-8601 Japan

14 3. Neurogenetics Branch, National Institute of Neurological Disorders and Stroke,

15 National Institutes of Health, Bethesda, MD, 20892 USA.

16 4. Biostatistics Section, Center for Advanced Medicine and Clinical Research,

17 Nagoya University Graduate School of Medicine, Nagoya, Japan

18 5. Department of Speech Pathology and Audiology, Aichi-Gakuin University School

1 of Health Science, 12 Araiike, Iwasaki-cho, Nisshin, 470-0195, Japan

2

3 Correspondence should be sent to Dr. Katsuno and Dr. Sobue:

4

5 Masahisa Katsuno, MD

6 Department of Neurology, Nagoya University Graduate School of Medicine,

7 65 Tsurumai-cho, showa-ku, Nagoya 466-8550 Japan

8 Tel: +81-52-744-2391; Fax: +81-52-744-2394; e-mail: ka2no@med.nagoya-u.ac.jp

9

10 Gen Sobue, MD

11 Department of Neurology, Nagoya University Graduate School of Medicine,

12 65 Tsurumai-cho, showa-ku, Nagoya 466-8550 Japan

13 Tel: +81-52-744-2385; Fax: +81-52-744-2384; e-mail: sobueg@med.nagoya-u.ac.jp

14

15 Word count: title, 98 characters; main text, 3520 words; Abstract: 198 words;

16 Introduction: 206 words; Number of figures and tables: 3 figures and 5 tables;

17 Number of references: 30 references

18 Supplemental data: 3 tables and 2 figures

1 **1. INTRODUCTION**

2 Spinal and bulbar muscular atrophy (SBMA), also known as Kennedy's disease, is an
3 adult-onset, hereditary neuromuscular disease characterized by extremity muscle
4 atrophy, weakness, contraction fasciculation, and bulbar involvement [1-3]. The
5 progression of neurological deficits is usually slow in SBMA, with the average
6 interval between the onset of symptoms and death approximately 20 years [4, 5].

7 There is currently no effective treatment to slow the progression of SBMA.
8 Although clinical trials of potential therapies have been done [6-9], definite efficacy
9 has not been clearly demonstrated in randomized controlled trials. These results
10 appear to be partly attributable to the absence of established outcome measures that
11 are sensitive to the disease-specific symptoms as well as the limited number of
12 patients, which may diminish statistical power. Namely, since the progression of
13 SBMA is slow, existing outcome measures such as those for amyotrophic lateral
14 sclerosis (ALS), are not sensitive enough to detect the deterioration of motor function
15 in the patients [10].

16 Thus, the present study aimed to develop and validate a disease-specific
17 functional rating scale for SBMA (SBMAFRS) and to evaluate whether this scale is
18 more advantageous than other existing outcome measures. In addition, for future use

1 in global clinical studies, the English version of SBMAFRS was also evaluated for

2 reliability.

3

1 **2. METHODS**

2 **2.1. Development of the SBMAFRS**

3 The SBMAFRS was designed as a quantitative outcome measure of global disability
4 in SBMA, especially for use in clinical trials. According to this concept, we developed
5 the SBMAFRS using the following four stages: item generation and scale
6 development; cross-cultural translation; scale validation; and scale evaluation.

7 At the item generation and scale development stage, we first listed up all the
8 items of the revised Amyotrophic Lateral Sclerosis Functional Rating Scale
9 (ALSFRS-R) and those of the modified Norris Scale, and created novel items
10 regarding bulbar and truncal function, which were not included in the existing scales.
11 Then seven experts of SBMA selected as many candidate items as possible with
12 consideration of the score distribution in our previous studies [7-10]. From the
13 candidates, the best suitable fourteen items were selected and modified to fit the
14 symptoms and severity of SBMA if necessary. Correspondence of the items of
15 SBMAFRS and those of the existing scales was shown in Table A-1. In the process of
16 determining item balance, we gave weight to disease-specific features and their
17 clinical relevance to SBMA. In the SBMAFRS, 5 of the 14 items addressed
18 bulbar-related symptoms because these are the primary symptoms of SBMA and are

1 occasionally lethal. Although pulmonary function tends to be preserved in SBMA,
2 one breathing-related item was adopted, because serious respiratory dysfunction has
3 been reported in certain subsets of patients and associated with the prognosis [5, 8].
4 This strategy may permit the identification of specific areas that respond to treatment
5 or the sub-classification of SBMA into etiologically or therapeutically relevant
6 subtypes.

7 We designed alternatives for each item so that the scale was sensitive enough
8 to detect subtle neurological changes. For example, although moderate dysphagia that
9 requires a change of dietary consistency is scored as 3 in the ALSFRS-R, we
10 subdivided this status into the following 2 scores in the SBMAFRS: score 2, avoids
11 some particular textures of food that may cause choking, etc.; and score 1, food
12 textures always have to be modified to a soft or chopped consistency. In addition, we
13 also ensured that the expressions used in the alternatives were as clear as possible to
14 avoid ambiguity. For instance, a specified frequency was required to choose the
15 alternative in the item for control of salivation: score 3, less than once a week; and
16 score 2, not less than once a week.

17 As a result, a functional rating scale consisting of 14 items was generated. At
18 the cross-cultural translation stage, we translated the original Japanese version

1 (SBMAFRS-J) into English (SBMAFRS-E) for the common use of this scale in
2 English-speaking countries. The translation procedure followed standard methods [11].
3 The SBMAFRS-J was translated into English considering cultural differences by a
4 bilingual translator, and it was later submitted to an expert committee. This committee,
5 consisting of C.G., G.S., and M.K., reviewed and revised the linguistic and conceptual
6 equivalence of the original Japanese version and the first English version. Next, this
7 English version was back translated into Japanese by a Japanese native without *a*
8 *priori* knowledge of the intent and concept of this scale. Finally, the expert committee
9 reviewed all reports and confirmed the linguistic and conceptual equivalence of this
10 scale, and completed the English version. At the scale validation stage, we assessed
11 the SBMAFRS-J on 80 subjects with SBMA, and evaluated its validity and reliability.
12 We also assessed the reliability of the SBMAFRS-E on 15 subjects with SBMA in the
13 USA. At the scale evaluation stage, we performed a 48-week longitudinal analysis on
14 an additional 41 Japanese subjects with SBMA to evaluate the usefulness of the
15 SBMAFRS-J as a potential endpoint in clinical trials. The evaluators of the
16 SBMAFRS were physicians or research nurses who were familiar with the medical
17 care for SBMA.

1 **2.2. Subjects and validation procedure**

2 We recruited a total of 80 consecutive male Japanese subjects with SBMA as the
3 SBMAFRS-J validation group. The inclusion criteria were as follows: 1) genetically
4 confirmed male Japanese SBMA subjects with more than one of the following
5 findings: muscle weakness, muscle atrophy, and bulbar dysfunction; 2) subjects who
6 were 25 to 75 years old at the time of informed consent; and 3) subjects who had not
7 undergone surgical castration.

8 To assess inter-rater reliability, a total of 80 Japanese participants were
9 assessed twice by two independent evaluators. To assess intra-rater reliability
10 (test-retest reproducibility), 26 of the 80 Japanese participants were assessed twice at
11 an interval of 1 to 3 weeks. This 1–3-week period was considered to be short enough
12 to avoid pathophysiological progression, but long enough to reduce the chance that
13 the subjects would recollect their previous responses.

14 We independently recruited a total of 41 male Japanese subjects with SBMA
15 as the SBMAFRS-J evaluation group and followed them for 48 weeks to assess the
16 usefulness of this outcome measure to detect disease progression. The inclusion
17 criteria were the same as those of the SBMAFRS-J validation group.

18 We also recruited a total of 15 male subjects with SBMA in the USA as the

1 SBMAFRS-E validation group. The inclusion criteria were the same as those of the
2 Japanese version. To assess inter-rater reliability, all participants were assessed twice
3 by two independent evaluators.

4 We followed the Japanese subjects at Nagoya University Hospital, and
5 collected their data between May 2010 and July 2014. The US subjects were followed
6 at NIH, and their data were collected between September 2012 and May 2013.

7 This study also included 41 Japanese male healthy controls matched for age and with
8 no significant medical history of neurological disorders. We informed the controls that
9 their results would be compared to those of subjects with SBMA, and they were
10 evaluated at Kojunkai Chuo Clinic, Tokai, Japan.

11 This study conformed to the ethics guidelines for human genome/gene
12 analysis research and to those for epidemiological studies endorsed by the Japanese
13 government. The ethical committee of Nagoya University Graduate School of
14 Medicine approved this study, and all participants gave their written informed consent.
15 All of the US subjects signed informed consent at the National Institutes of Health in
16 Bethesda, Maryland, USA.

17

18 **2.3. Other clinical parameters**

We used two other functional rating scales that are commonly applied to evaluate disease severity, i.e., the ALSFRS-R and the modified Norris score, to assess the external validity of the SBMAFRS.

The ALSFRS is a validated, questionnaire-based scale that measures the physical function of subjects with ALS in carrying out activities of daily living [12].

The revised version of this scale, the ALSFRS-R, was generated to improve the disproportionate weighting to limb and bulbar function, as compared to respiratory dysfunction. Each of the 12 items is rated in 4 ordinal categories, and thus the best possible score is 48. The ALSFRS-R was translated into Japanese, and its validity has been confirmed [13].

The modified Norris Score is another rating scale for ALS that consists of two parts, i.e., the Limb Norris Score and the Norris Bulbar Score. The Limb Norris Score has 21 items to evaluate limb function and the Norris Bulbar Score has 13 items to assess bulbar function. Each item is rated in 4 ordinal categories, and thus the best possible scores are 63 and 39, respectively. The original version was translated into Japanese, and its validity has been confirmed [14].

2.4. Statistical analysis

1 Statistical analyses were performed using SPSS Statistics 17.0 (SPSS Japan, Inc.,
2 Tokyo, Japan). We used descriptive variables such as the mean and standard deviation
3 to summarize the quantitative measures. The normality of the score distribution was
4 assessed by Shapiro-Wilk test. The score distributions were also assessed by skewness
5 and kurtosis. For multiple comparisons, *p* values were corrected using Dunnett's test.
6 We examined scaling assumption (confirmatory factor analyses); acceptability
7 (distribution of the subscales and total score); reliability (test-retest reliability:
8 intra-class correlation coefficients [ICCs] with a one-way mixed effect model;
9 inter-rater reliability: ICCs with a two-way mixed effect model); and validity (internal
10 consistency: Cronbach's alpha; inter-scale correlations; and convergent and
11 discriminate construct validity). Often, 0.70 is recommended as a minimum standard
12 for ICCs to indicate reliability. A value of 0.70–0.90 is considered to indicate good
13 internal consistency for Cronbach's alpha [15]. In the longitudinal analysis, a
14 standardized response mean (SRM) was also calculated as an index of the effect size
15 for a direct comparison among the outcome measures. Values of 0.20, 0.50, and 0.80
16 were considered to represent small, moderate, and large changes over time,
17 respectively [16]. The SRM was calculated using the ratio of the mean score change
18 to the standard deviation of score change.

1 **3. RESULTS**

2 **3.1. Items and alternatives of the SBMAFRS**

3 The SBMAFRS consists of 14 items, each of which contains 5 (0–4) alternatives
4 (Table 1). The 14 items consist of 5 subscales: bulbar-, upper limb-, trunk-, lower
5 limb-, and breathing-related. The total available score ranges from 0 (worst) to 56
6 (normal). The bulbar-related subscale contains 5 items, the upper limb-related
7 subscale contains 2 items, the trunk-related subscale contains 4 items, the lower
8 limb-related subscale contains 2 items, and the breathing-related subscale contains 1
9 item. To minimize the difference in the questioning and examination styles used
10 among raters, a rating algorithm that specified the exact procedures to be used was
11 attached to this scale (Fig. A-1).

12

13 **3.2. Demographics of study population**

14 The background characteristics of the subjects included in the present study were
15 similar to those of previous clinical studies [7, 8, 10]. There was no significant
16 difference in demographics among the three study populations: Scale validation group
17 (Japan), Scale evaluation group (Japan), and Scale validation group (USA) (Table 2).

18

3.3. Score distribution

The scores of the SBMAFRS-J and SBMAFRS-E were 36.5 ± 7.4 (20–54) and 37.3 ± 3.8 (31–43), respectively. The total score was distributed in a normal manner (Fig. 1A). Absolute values of skewness and kurtosis of the SBMAFRS total scores were smaller than those of other functional scales, suggesting the distribution normality of SBMAFRS (Table A-2). The comparison between the total score distribution of the SBMAFRS-J and ALSFRS-R showed that the SBMAFRS-J scores varied from low to high without a ceiling or floor effect, although the distribution of the ALSFRS-R scores was skewed to the right (Fig. 1B). The total score of the SBMAFRS-J for the 41 healthy controls was 55.9 ± 0.4 (54–56), which was close to the highest possible score.

12

3.4. Scale validation

3.4.1. Reliability estimates

Intra-rater agreement based on the total score of the SBMAFRS-J obtained in 26 Japanese subjects with SBMA was excellent. Detailed analysis of the ICCs for each subscale demonstrated that all subscales, except for the breathing function subscale, had an excellent intra-rater agreement (Table 3). Inter-rater reliability was also high,

1 although detailed analysis of the ICCs for each item indicated lower values than for
2 intra-rater reliability. Internal consistency was satisfactory, with Cronbach's alpha
3 values ranging from 0.700 to 0.822 (Table 3). Inter-rater reliability and internal
4 consistency of the SBMAFRS-E was similar to those of the Japanese version of this
5 scale, SBMAFRS-J, although the ICC for the breathing-related subscale of the
6 SBMAFRS-E was better than that of the SBMAFRS-J.

7

8 3.4.2. *Validity estimate*

9 As a “gold standard” disease-specific functional parameter has not been established
10 for SBMA, we could not evaluate the criterion validity. Instead, we assessed the
11 correlation of the subscales of the SBMAFRS with those of existing functional scales
12 to evaluate their convergent and discriminant construct validity (Table 4). The
13 directions and patterns of correlations were consistent with what we assumed. For
14 example, the bulbar-related subscale of the SBMAFRS correlated best with that of the
15 ALSFRS-R and Norris Bulbar Score, and worst with the lower limb- or
16 respiration-related subscale of the ALSFRS-R and Limb Norris Score. Conversely, the
17 lower limb-related subscale of the SBMAFRS correlated best with that of the
18 ALSFRS-R and Limb Norris Score, and worst with the bulbar-related subscale of the

1 ALSFRS-R and Norris Bulbar Score.

2

3 *3.4.3. Confirmatory factor analysis*

4 As stated above, the SBMAFRS was developed with the assumption that the items
5 can be categorized into the following 5 subscales: bulbar-, upper limb-, trunk-, lower
6 limb-, and breathing-related. To verify this assumption, we applied factor analysis
7 using oblique varimax rotation for the SBMAFRS-J (Table A-3). This analysis
8 extracted 3 factors that contributed 60.0% of the total variance. Factor 1 mainly
9 included the trunk- and lower limb-related subscales, while the bulbar- and upper
10 limb-related subscales both contributed to factor 2. These results confirmed the
11 factorial validity of the SBMAFRS, although the bulbar and upper limb subscales
12 were not divided; neither were the trunk and lower limb subscales. Thus, the motor
13 function of the SBMA subjects assessed using the SBMAFRS appears to be classified
14 into 3 domains: bulbar and upper limb; trunk and lower limb; and respiratory. These
15 findings were consistent with our previous study showing that bulbar impairment is
16 closely related to upper limb dysfunction, and trunk dysfunction is associated with
17 lower limb weakness in SBMA [10].

18

1 *3.4.4. Longitudinal analysis in the scale evaluation group*

2 To evaluate the sensitivity of the SBMAFRS as an outcome measure, we
3 prospectively analyzed longitudinal changes in the SBMAFRS-J and other functional
4 parameters for 48 weeks. The results showed slow but steady disease progression in
5 all subscales of the SBMAFRS-J, as well as in the other functional parameters (Table
6 5, Fig. 2, and Fig. A-3). To facilitate a direct comparison of responsiveness among the
7 various functional parameters, we calculated the SRM as an index of the effect size by
8 the longitudinal mean change/standard deviation of score change for each parameter
9 (Table 5). The SRM, which enables a comparison of sensitivity among the outcome
10 measures, of the SBMAFRS was larger than that of the other scales, indicating that
11 the SBMAFRS would require the smallest sample size in clinical trials and thus
12 appears to be the most sensitive functional parameter [17, 18]. Sample size estimation
13 based on this longitudinal analysis was the lowest for the SBMAFRS, followed by the
14 ALSFRS-R and Norris Bulbar Score, suggesting that the SBMAFRS was a sensitive
15 clinical measure which detects the disease progression over time (Fig. 3).

1 **4. DISCUSSION**

2 Outcome measures that enable the rigorous quantification of disease severity or
3 quality of life are essential to estimate the efficacy of potential therapies [19-24].
4 Disease-specific functional rating scales that measure global dysfunction, based on a
5 structured psychometric validation procedure, are increasingly important in defining
6 the primary endpoint in clinical trials, because the quality of the rating scales has the
7 potential to determine the validity of such studies [25-27]. Nevertheless, to date, no
8 disease-specific outcome measures for SBMA are available. In the present study, we
9 created a disease-specific functional rating scale for SBMA (SBMAFRS) and
10 validated this score in cross-sectional and longitudinal analysis. One major advantage
11 of the SBMAFRS is that it was developed as a compound scale to capture the multiple
12 aspects of SBMA. The results of the present study suggest that the SBMAFRS were
13 valid and reliable measure for the quantitative evaluation of SBMA. In the
14 longitudinal evaluation, the change of the SBMAFRS-J over time was more readily
15 detectable than that of the other functional parameters, although still larger sample
16 sizes are required for detection of therapeutic efficacy. These results demonstrate that
17 the SBMAFRS is a disease-specific outcome measure that can be used for clinical
18 studies.

1 The SBMAFRS was designed to overcome several problems that arose
2 when we applied the functional measures for ALS or other neuromuscular disorders to
3 SBMA. First, it tends to be difficult to notice the chronological changes of
4 neurological symptoms in subjects with SBMA because disease progression in this
5 disease is much slower than in ALS. In the process of scale development, each item
6 was designed to detect subtle symptoms by subdividing the alternatives of the milder
7 symptoms. Second, the vague descriptions of the alternatives in the existing clinical
8 measures might result in different responses between raters. We developed each
9 alternative of the SBMAFRS to ensure that the expressions used were as clear as
10 possible. As a result, the scores of the SBMAFRS were distributed in a wider range
11 than those of the ALSFRS-R, suggesting that the SBMAFRS measures the physical
12 function of individual subjects with mild motor dysfunction more sensitively than the
13 ALSFRS-R.

14 The confirmatory factor analysis of this scale indicated that the bulbar
15 function is strongly associated with the upper limb function in spinal and bulbar
16 muscular atrophy, as we previously demonstrated [10]. Similar relationship is also
17 shown in ALS: the degree of bulbar palsy is more strongly associated with upper limb
18 weakness than lower limb weakness [28]. As a possible pathomechanism for this

phenomenon, spread of lesions from the bulbar region to the upper limb is demonstrated in ALS [29]. Given that the spread of neuropathology is also suggested in polyglutamine diseases, a similar mechanism may underlie the pathophysiology of SBMA [30].

The SBMAFRS-J was also examined prospectively over 48 weeks to evaluate its sensitivity to changes over time. To this end, we calculated the SRM as the index of the effect size. Calculation of the SRM enables a direct comparison of the sensitivity of the scales to detect changes among outcome measures. Compared to the other measures, the SBMAFRS had a larger SRM, indicating that its use might reduce the number of subjects needed for clinical trials. The effect sizes of disease-modifying therapies in slowly progressive neurodegenerative disorders generally tend to be small because of smaller longitudinal changes in proportion to large symptomatic daily fluctuations [31-33]. Although a relatively small SRM is an inevitable feature of slowly progressive neurodegenerative diseases such as SBMA, effort should be made to further increase the sensitivity of this scale through the experience gained by its use in future clinical trials.

Although our results suggest that the SBMAFRS is a valid disease-specific score of SBMA, there are limitations that need to be improved. First, low reliability of

1 the breathing-related item is a critical drawback of the SBMAFRS-J, which is
2 attributable to several factors. One of the most important issues was that each
3 alternative of breathing-related item was categorized within a narrow range, because
4 respiratory function in the most of patients with SBMA is preserved and a few need
5 ventilation support [5]. This over-categorization might lead to the discrepancies
6 among evaluations and result in low intra-rater or inter-rater agreement. In addition, a
7 minor disagreement led to low ICC for the breathing-related item of the SBMAFRS-J,
8 since the score distribution biased extremely: 58 out of 80 subjects were judged as
9 having the same score in the inter-rater agreement assessment (Table A-4A). Similar
10 results were also obtained as for the intra-rater reproducibility (Table A-4B).
11 Furthermore, the number of raters was eight for the SBMAFRS-J, but two for
12 SBMAFRS-E. This difference might also lead to the differential reliability of
13 breathing-related subscale of the SBMAFRS-E, since only one item was included in
14 this subscore. The small number of the US subjects and relatively short longitudinal
15 follow-up period also limit the interpretation of our results. In the future study, a
16 larger cohort needs to be followed-up for longer time to confirm the utility of the
17 SBMAFRS. Second, the old-fashioned clinimetrics that adopted items without
18 standardization or weighing as in SBMAFRS would be disadvantageous for being

used in clinical trials because of the limited linearity. In this regard, the Rasch and items response theory is an important approach to transform ordinal scores into interval measures that are scale-independent and accurate for patients' assessment [34]. In fact, this methodology has been applied for several functional scales of neurological disorders such as the 9-item fatigue severity scale (FSS) for immune-mediated neuropathies [35] and the modified Medical Research Council grading system [36]. Furthermore, the concept of minimal clinically important difference (MCID) should also be considered as a means of assessing treatment response in clinical trials [37].

In summary, we created a disease-specific functional rating scale for SBMA, a slowly progressive neurodegenerative disorder. Although our results indicate the validity and sensitiveness of this scale, the SBMAFRS, the feasibility of this scale as an outcome measure needs to be re-assessed in interventional studies with a larger sample size including the English-speaking subjects.

1 **Funding**

2 Dr. Katsuno is supported by a Grant-in-Aid for Scientific Research on Innovated
3 Areas “Foundation of Synapse and Neurocircuit Pathology” from MEXT, Japan
4 (22110005); KAKENHI grants from MEXT/JSPS, Japan (No. 22110005, 26293206,
5 26670440, and 26670439); and Core Research for Evolutional Science and
6 Technology (CREST) from the Japan Science and Technology Agency (JST); and a
7 grant from the Daiichi Sankyo Foundation of Life Science.

8 Dr. Sobue is supported by KAKENHI grants from MEXT/JSPS, Japan (No.
9 21229011); grants from the Ministry of Welfare, Health and Labor of Japan; and Core
10 Research for Evolutional Science and Technology (CREST) from the Japan Science
11 and Technology Agency (JST).

12

13

1 **References**

- 2 [1] Kennedy WR, Alter M, Sung JH. Progressive proximal spinal and bulbar muscular
3 atrophy of late onset. A sex-linked recessive trait. *Neurology* 1968;18:671-680.
- 4 [2] Sobue G, Hashizume Y, Mukai E, et al. X-linked recessive bulbospinal
5 neuronopathy. A clinicopathological study. *Brain* 1989;112:209-232.
- 6 [3] Katsuno M, Tanaka F, Adachi H et al. Pathogenesis and therapy of spinal and
7 bulbar muscular atrophy (SBMA). *Prog Neurobiol.* 2012; 99(3):246-256.
- 8 [4] Sperfeld AD, Karitzky J, Brummer D et al. X-linked bulbospinal neuronopathy:
9 Kennedy disease. *Arch Neurol* 2002;59:1921-1926.
- 10 [5] Atsuta N, Watanabe H, Ito M et al. Natural history of spinal and bulbar muscular
11 atrophy (SBMA): a study of 223 Japanese patients. *Brain.* 2006;129:1446-1455.
- 12 [6] Banno H, Katsuno M, Suzuki K et al. Phase 2 trial of leuprorelin in patients with
13 spinal and bulbar muscular atrophy. *Ann Neurol* 2009;65:140-150.
- 14 [7] Katsuno M, Banno H, Suzuki K et al. Efficacy and safety of leuprorelin in patients
15 with spinal and bulbar muscular atrophy (JASMITT study): a multicentre,
16 randomised, double-blind, placebo-controlled trial. *Lancet Neurol* 2010;9:875-884.
- 17 [8] Fernández-Rhodes LE, Kokkinis AD, White MJ et al. Efficacy and safety of
18 dutasteride in patients with spinal and bulbar muscular atrophy: a randomised

- 1 placebo-controlled trial. *Lancet Neurol* 2011;10:140-147.
- 2 [9] Querin G, D'Ascenzo C, Peterle E et al. Pilot trial of clenbuterol in spinal and
3 bulbar muscular atrophy. *Neurology* 2013;80:2095-2098.
- 4 [10] Hashizume A, Katsuno M, Banno H et al. Longitudinal changes of outcome
5 measures in spinal and bulbar muscular atrophy. *Brain* 2012;135:2838-2848.
- 6 [11] Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health related
7 quality of life measures: literature review and proposed guidelines. *J Clin*
8 *Epidemiol* 1993;46:1417-1432.
- 9 [12] The ALS CNTF treatment study (ACTS) phase I-II Study Group. The
10 Amyotrophic Lateral Sclerosis Functional Rating Scale. Assessment of activities of
11 daily living in patients with amyotrophic lateral sclerosis. *Arch Neurol*
12 1996;53:141-147.
- 13 [13] Ohashi Y, Tashiro K, Itoyama Y et al. Study of functional rating scale for
14 amyotrophic lateral sclerosis: revised ALSFRS(ALSFRS-R) Japanese version. *No*
15 *To Shinkei* 2001;53:346-355.
- 16 [14] Oda E, Ohashi Y, Tashiro K, et al. Reliability and factorial structure of a rating
17 scale for amyotrophic lateral sclerosis. *No To Shinkei*. 1996;48:999-1007.
- 18 [15] Nunnally JC, Bernstein IH. *Psychometric theory*. 3rd ed. New York:

- 1 McGraw-Hill; 1994.
- 2 [16] Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in
3 health status. *Med Care* 1989;27(3 Suppl.):S178-S189.
- 4 [17] Giordano A, Pucci E, Naldi P et al. Responsiveness of patient reported outcome
5 measures in multiple sclerosis relapses: the REMS study. *J Neurol Neurosurg*
6 *Psychiatry*. 2009;80:1023-1028.
- 7 [18] Schmitz-Hübsch T, Fimmers R, Rakowicz M et al. Responsiveness of different
8 rating instruments in spinocerebellar ataxia patients. *Neurology*. 2010;74:678-684.
- 9 [19] Burns J, Ouvrier R, Estilow T et al. Validation of the Carcot-Marie-Tooth Disease
10 Pediatric Scale as an Outcome Measure of Disability. *Ann Neurol*
11 2012;71:642-652.
- 12 [20] Simmons Z, Felgoise SH, Bremer BA et al. The ALSSQOL. Balancing physical
13 and nonphysical factors in assessing quality of life in ALS. *Neurology*
14 2006;67:1659-1664.
- 15 [21] Shy ME, Blake J, Krajewski K et al. Reliability and validation of the CMT
16 neuropathy score as a measure of disability. *Neurology* 2005;64:1209-1214.
- 17 [22] Müller J, Wissel J, Kemmler G et al. Craniocervical dystonia questionnaire
18 (CDQ-24): development and validation of a disease-specific quality of life

- 1 instrument. J Neurol Neurosurg Psychiatry 2004;75:749-753.
- 2 [23] Payan CA, Viallet F, Landwehrmeyer BG et al. Disease severity and progression
3 in progressive supranuclear palsy and multiple system atrophy: validation of the
4 NNIPPS--Parkinson Plus Scale. PLoS One. 2011;6(8):e22293.
- 5 [24] Hashizume A, Katsuno M, Banno H et al. Difference in chronological changes of
6 outcome measures between untreated and placebo-treated patients of spinal and
7 bulbar muscular atrophy. J Neurol. 2012;259:712-719
- 8 [25] Huntington Study Group. Unified Huntington's Disease Rating Scale: Reliability
9 and-Consistency. Movement Disord 1996;11,136-142.
- 10 [26] Wenning GK, Tison F, Seppi K et al. Development and Validation of the
11 Unified Multiple System Atrophy Rating Scale (UMSARS). Movement Disord
12 2004;19:1391-1402.
- 13 [27] Hobart JC, Cano SJ, Zajicek JP, et al. Rating scales as outcome measures for
14 clinical trials in neurology: problems, solutions, and recommendations. Lancet
15 Neurol 2007;6:1094-1105.
- 16 [28] Simon NG, Lomen-Hoerth C, Kiernan MC. Patterns of clinical and
17 electrodiagnostic abnormality in early amyotrophic sclerosis. Muscle Nerve. 2014;
18 50: 894-9.

- 1 [29] Ravits JM, La Spada AR. ALS motor phenotype heterogeneity, focality, and
2 spread: deconstructing motor neuron degeneration. *Neurology*. 2009; 73: 805-811.
- 3 [30] Cicchetti F, Lacroix S, Cisbani G, et al., Mutant huntingtin is present in neuronal
4 grafts in Huntington disease patients. *Ann Neurol*. 2014; 76: 31-42.
- 5 [31] Marder K, Zhao H, Myers RH et al. Rate of functional decline in Huntington's
6 disease. Huntington Study Group. *Neurology*. 2000;54:452-458.
- 7 [32] Martinez-Martin P, Prieto L, Forjaz MJ. Longitudinal metric properties of
8 disability rating scales for Parkinson's disease. *Value Health*. 2006;9:386-393.
- 9 [33] Martínez-Martin P, Carod-Artal FJ, da Silveira Ribeiro L et al. Longitudinal
10 psychometric attributes, responsiveness, and importance of change: An approach
11 using the SCOPA-Psychosocial questionnaire. *Mov Disord*. 2008;23:1516-1523.
- 12 [34] Merkies IS, Faber CG. Outcome measures in Duchenne muscular dystrophy: are
13 we ready for the new therapeutic era? *Neuromuscul Disord*. 2009;19:447.
- 14 [35] Draak TH, Vanhoutte EK, van Nes SI, et al., Changing outcome in inflammatory
15 neuropathies: Rasch-comparative responsiveness. *Neurology*. 2014; 83: 2124-2132.
- 16 [36] Vanhoutte EK, Faber CG, van Nes SI et al., Modifying the Medical Research
17 Council grading system through Rasch analyses. *Brain*. 2012;135:1639-49.
- 18 [37] Smith AG, Burns TM. Reevaluating clinical measurement tools in therapeutic

1 trials: time to make a Rasch decision? Neurology. 2014;83:2104-2105.

2

1 **Figure legends**

2

3 **Fig. 1. Score distribution of the SBMAFRS and the ALSFRS-R.** The SBMAFRS-J
4 distributed in a normal manner (A). The total score distribution of the SBMAFRS varied from low
5 to high without ceiling or floor effects (A), although that of the ALSFRS-R skewed to the right
6 with a ceiling effect (B).

7

8 **Fig. 2. Individual change of the functional scores.** The change of the total score of
9 SBMAFRS (A) and that of ALSFRS-R (B) in individual subjects.

10

11 **Fig. 3. Sample size estimation.** Estimated sample size per arm for a 2-arm
12 interventional clinical trial based on the observed variability of the chronological
13 progression rate found in the present study (power 0.80, α 0.05). SBMAFRS =
14 Spinal and Bulbar Muscular Atrophy Functional Rating Scale; ALSFRS-R = Revised
15 Amyotrophic Lateral Sclerosis Functional Rating Scale

16

Table 1. Spinal and bulbar muscular atrophy functional rating scale (SBMAFRS)

Bulbar-related subscale	
Speech	<p>4. Normal speech.</p> <p>3. Speaks in a slightly nasal tone.</p> <p>2. What is said can be understandable despite impaired speech.</p> <p>1. A portion of what is said cannot be understandable unless heard over and over again.</p> <p>0. Other media than voices are needed for communication.</p>
Control of salivation	<p>4. No drooling.</p> <p>3. Dribbles at night or during eating (less often than once a week).</p> <p>2. Occasionally dribbles at night or during eating (not less often than once a week).</p> <p>1. Dribbles on other occasions than above.</p> <p>0. Dribbles so much that always needs tissues or a handkerchief.</p>
Swallowing	<p>4. No difficulty.</p> <p>3. Occasionally chokes on food depending on its texture.</p> <p>2. Avoids some particular textures of food that may cause choking, etc.</p> <p>1. Food textures have to be always modified to soft or chopped ones.</p> <p>0. Enteral or intravenous nutrition is needed.</p>
Tongue	<p>4. No atrophy.</p> <p>3. Atrophic, but the tongue tip can be smoothly moved to and fro between the left and right corners of the mouth.</p> <p>2. The tongue tip can be moved, but not smoothly, to and fro between the left and right corners of the mouth.</p> <p>1. The tongue tip cannot reach the mouth corner.</p> <p>0. No motion of the tongue.</p>
Puffing cheeks	<p>4. The air does not leak when the cheeks are firmly pressed.</p> <p>3. The air leaks when the cheeks are firmly pressed.</p> <p>2. The cheeks can be puffed out, but not fully.</p> <p>1. Although the lips can be pursed, the cheeks cannot be puffed.</p> <p>0. The lips cannot be pursed.</p>

Upper limb-related subscale	
Writing	<p>4. No difficulty.</p> <p>3. Slightly difficult in writing, but can write a long sentence as well.</p> <p>2. Writing gets slow in writing a long sentence.</p> <p>1. Difficult in writing even a short sentence.</p> <p>0. Unable to write letters.</p>
Eating action	<p>4. No problem with eating actions.</p> <p>3. Able to finish eating within a normal time, but has some difficulty in using a knife and a fork.</p> <p>2. Takes a longer eating time due to his difficulty in using a knife and a fork.</p> <p>1. Unable to use a knife or a fork, but good enough at using a spoon.</p> <p>0. Needs assistance for eating.</p>
Trunk-related subscale	
Dressing activity	<p>4. No problem.</p> <p>3. Partially impaired, but taking no additional time to change clothes.</p> <p>2. Partially impaired, so takes time to change clothes because he needs to sit or lean against something.</p> <p>1. Partial assistance is needed.</p> <p>0. Total assistance is needed.</p>
Rising from a sitting position	<p>4. No problem.</p> <p>3. Minor problem.</p> <p>2. Able to rise all alone, but needs to hold on something.</p> <p>1. Any assistance is needed.</p> <p>0. Unable to rise at all.</p>
Arising from a supine position	<p>4. No problem.</p> <p>3. Minor problem.</p> <p>2. Able to arise all alone, but needs some efforts, ex.) holding a lateral position beforehand.</p> <p>1. Any assistance is needed to arise.</p> <p>0. Unable to arise at all.</p>

Bowing	<p>4. Able to bow his head fully in a standing position.</p> <p>3. Able to bow his head in a standing position, but not fully.</p> <p>2. Able to hold a standing position, but slightly hard to bow his head.</p> <p>1. Difficult in holding a standing position, but able to bow his head in a sitting position.</p> <p>0. Unable to bow his head even in a sitting position.</p>
Lower limb-related subscale	
Walking	<p>4. No problem with walking or running.</p> <p>3. No problem with walking, but it is difficult to run.</p> <p>2. A cane or some other assistive device is sometimes needed.</p> <p>1. A cane or some other assistive device is always needed.</p> <p>0. Walking is impossible.</p>
Stairs	<p>4. No problem with walking upstairs.</p> <p>3. Able to walk upstairs without a handrail, but slowly.</p> <p>2. Able to walk upstairs without assistance if a handrail is available.</p> <p>1. Assistance from someone is needed.</p> <p>0. Unable to walk upstairs anyhow.</p>
Breathing-related subscale	
Breathing	<p>4. No dyspnea</p> <p>3. Gets short of breath during activity like walking.</p> <p>2. Gets short of breath even during slight activity like standing up.</p> <p>1. Has shortness of breath even at rest.</p> <p>0. A respiratory assistor such as NIPPV and TPPV is needed.</p>

Table 2. Clinical profiles of the Japanese and US patients with SBMA

	Scale validation group (Japan) (n = 80)	Scale evaluation group (Japan) (n = 41)	Scale validation group (USA) (n = 15)	<i>P</i>
Age at first evaluation (years)	57.7 ± 10.0 (37 - 82)	51.9 ± 11.2 (30 - 73)	56.1 ± 10.9 (32 -46)	NS
Disease duration (years)	14.7 ± 7.3 (1 - 37)	9.8 ± 5.7 (1 - 24)	17.5 ± 12.2 (3 - 49)	NS
Age at onset (years)	43.1 ± 11.4 (18 - 66)	41.7 ± 11.8 (22 - 66)	38.6 ± 13.5 (18 - 60)	NS
<u>CAG repeat length on AR gene</u>	<u>48.3 ± 3.7</u> <u>(40 - 60)</u>	<u>49.6 ± 4.8</u> <u>(40 - 60)</u>	<u>46.7 ± 2.6</u> <u>(43 - 51)</u>	<u>NS</u>
SBMAFRS total score	36.5 ± 7.4 (20 - 54)	41.3 ± 6.9 (25 - 51)	37.3 ± 3.8 (31 - 43)	NS
ALSFRS-R total score	39.1 ± 4.2 (27 - 47)	41.8 ± 3.4 (34 - 48)	37.7 ± 3.2 (32 - 46)	NS
Creatine kinase (IU/L)	781 ± 653 (103 - 4164)	1189 ± 763 (249 - 2296)	1119 ± 512 (491 - 2296)	NS
Creatinine (mg/dL)	0.41 ± 0.13 (0.19 - 0.79)	0.46 ± 0.08 (0.30 - 0.72)	0.51 ± 0.11 (0.33 - 0.72)	NS
Testosterone (µg/dL)	5.12 ± 3.82 (0.20 - 18.42)	7.01 ± 2.80 (2.77 - 13.82)	4.38 ± 1.58 (1.47 - 7.34)	NS

SBMAFRS = spinal and bulbar muscular atrophy functional rating scale; ALSFRS-R = revised amyotrophic lateral sclerosis functional rating scale;
NS = not significant. Each cell contains the mean ± SD.

Table 3. Intra- and inter-rater reliability and internal consistency of each subscale in the SBMAFRS

		Bulbar	Upper limbs	Trunk	Lower limbs	Breathing	Total score
Reliability							
Intra-rater agreement (95% CI)	(Japanese) (n = 26)	0.890 (0.772–0.949)	0.828 (0.656–0.919)	0.935 (0.863–0.970)	0.957 (0.907–0.980)	0.076 (-0.309–0.442)	0.910 (0.811–0.958)
	(Japanese) (n = 80)	0.658 (0.477–0.781)	0.745 (0.602–0.836)	0.817 (0.715–0.883)	0.831 (0.737–0.892)	0.297 (0–0.549)	0.797 (0.684–0.870)
Inter-rater agreement (95% CI)	(US) (n = 15)	0.808 (0.428–0.936)	0.665 (0.004–0.888)	0.786 (0.361–0.928)	0.823 (0.473–0.941)	0.985 (0.955–0.995)	0.837 (0.515–0.945)
Internal consistency							
Cronbach's alpha	(Japanese) (n = 80)	0.700	0.712	0.822	0.784	NA	NA
	(US) (n = 15)	0.438	0.393	0.519	0.494	NA	NA

NA = not applicable; CI = confidence interval. Inter-rater and intra-rater agreement were evaluated by using the intra-class correlation coefficient.

Table 4. Convergent and discriminant construct validity of the SBMAFRS-J

		SBMAFRS-J					
		Bulbar	Upper limbs	Trunk	Lower limbs	Breathing	Total score
ALSFRS-R	Bulbar	0.873 <i>p</i> < 0.001	0.542 <i>p</i> < 0.001	0.257 <i>p</i> = 0.042	0.196 <i>p</i> = 0.124	0.222 <i>p</i> = 0.080	0.635 <i>p</i> < 0.001
	Upper limbs	0.549 <i>p</i> < 0.001	0.856 <i>p</i> < 0.001	0.469 <i>p</i> < 0.001	0.432 <i>p</i> < 0.001	0.355 <i>p</i> = 0.004	0.714 <i>p</i> < 0.001
	Trunk	0.354 <i>p</i> = 0.004	0.436 <i>p</i> < 0.001	0.741 <i>p</i> < 0.001	0.747 <i>p</i> < 0.001	0.279 <i>p</i> = 0.027	0.734 <i>p</i> < 0.001
	Lower limbs	0.186 <i>p</i> = 0.145	0.385 <i>p</i> = 0.002	0.834 <i>p</i> < 0.001	0.852 <i>p</i> < 0.001	0.312 <i>p</i> = 0.013	0.693 <i>p</i> < 0.001
	Respiration	0.043 <i>p</i> = 0.736	0.115 <i>p</i> = 0.367	0.226 <i>p</i> = 0.075	0.268 <i>p</i> = 0.034	0.662 <i>p</i> < 0.001	0.247 <i>p</i> = 0.051
Limb Norris score		0.370 <i>p</i> = 0.003	0.578 <i>p</i> < 0.001	0.847 <i>p</i> < 0.001	0.849 <i>p</i> < 0.001	0.366 <i>p</i> = 0.003	0.853 <i>p</i> < 0.001
Norris Bulbar score		0.801 <i>p</i> < 0.001	0.511 <i>p</i> < 0.001	0.365 <i>p</i> = 0.003	0.292 <i>p</i> = 0.020	0.181 <i>p</i> = 0.156	0.680 <i>p</i> < 0.001
Age at examination		0.019 <i>p</i> = 0.879	0.032 <i>p</i> = 0.797	-0.269 <i>p</i> = 0.027	-0.179 <i>p</i> = 0.144	-0.161 <i>p</i> = 0.189	-0.145 <i>p</i> = 0.238
Disease duration		0.100 <i>p</i> = 0.419	-0.140 <i>p</i> = 0.255	-0.347 <i>p</i> = 0.004	-0.427 <i>p</i> < 0.001	-0.095 <i>p</i> = 0.442	-0.240 <i>p</i> = 0.049

SBMAFRS = spinal and bulbar muscular atrophy functional rating scale; ALSFRS-R = revised amyotrophic lateral sclerosis functional rating scale. Each cell contains Spearman's correlation coefficient (upper line) and the corresponding *p* value (lower line).

Table 5. Longitudinal analyses of the SBMAFRS-J and other functional parameters

N = 41	Baseline (SD)	Follow-up (SD)	Mean longitudinal change (SE)	SRM
SBMAFRS				
Bulbar	14.5 (3.1)	14.4 (2.9)	-0.22 (0.25)	-0.14
Upper limbs	6.0 (1.7)	5.7 (1.6)	-0.27 (0.15)	-0.29
Trunk	12.0 (2.5)	11.3 (2.6)	-0.66 (0.22)	-0.49
Lower limbs	5.0 (1.4)	4.9 (1.2)	-0.10 (0.15)	-0.11
Breathing	3.8 (0.4)	3.7 (0.5)	-0.02 (0.09)	-0.05
Total score	41.3 (6.9)	40.0 (6.2)	-1.27 (0.44)	-0.48
ALSFRS-R	41.8 (3.4)	41.0 (3.2)	-0.73 (0.31)	-0.39
Limb Norris Score	55.7 (5.3)	55.1 (5.3)	-0.56 (0.38)	-0.25
Norris Bulbar Score	33.3 (4.8)	32.6 (4.7)	-0.73 (0.42)	-0.29

SBMAFRS = spinal and bulbar muscular atrophy functional rating scale; ALSFRS-R = revised amyotrophic lateral sclerosis functional rating scale; SD = standard deviation; SE = standard error of the mean; SRM = standardized response mean.

Figure 1

(A) Total score of SBMAFRS

(B) Total score of ALSFRS-R

Figure 2

(A) SBMAFRS

(B) ALSFRS-R

Figure 3