Published Ahead of Print on June 19, 2023 as 10.1212/WNL.0000000000207483

OPEN

Neurology®

The most widely read and highly cited peer-reviewed neurology journal
The Official Journal of the American Academy of Neurology

AMERICAN ACADEMY OF NEUROLOGY®

Neurology Publish Ahead of Print
DOI:10.1212/WNL.0000000000207483

Development of a Rasch-Built Amyotrophic Lateral Sclerosis Impairment Multidomain Scale to Measure Disease Progression in ALS

Author(s):
Adriaan D. de Jongh, MD PhD MSc[1,2]; Ruben P.A. van Eijk, M.D., PhD.[1,3]; Leonhard A. Bakker, MSc[1,4]; Tommy M. Bunte, MSc[1]; Anita Beelen, PhD[4,5]; Conny van der Meijden[6]; Michael A. van Es, M.D., PhD.[1]; Johanna M.A. Visser-Meily, PhD, Professor[4,5]; Esther T. Kruitwagen, MD PhD[4,5]; Jan H Veldink, MD, PhD[1]; Leonard H. van den Berg, M.D., PhD.[1]

Corresponding Author:
Leonard H. van den Berg, lberg@umcutrecht.nl

Affiliation Information for All Authors: 1. Department of Neurology, UMC Utrecht Brain Center, University Medical Center Utrecht, Utrecht, the Netherlands; 2. Department of Neurology, Tergooi hospital, Blaricum, the Netherlands; 3. Biostatistics & Research Support, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, the Netherlands; 4. Department of Rehabilitation, Physical Therapy Science and Sports, UMC Utrecht Brain Center, University Medical Center Utrecht, Utrecht, the Netherlands; 5. Center of Excellence for Rehabilitation Medicine, UMC Utrecht Brain Center, University Medical Center Utrecht, and De Hoogstraat Rehabilitation, Utrecht, the Netherlands; 6. ALS Patients Connected, Bilthoven, the Netherlands

Equal Author Contribution:

Contributions:
Adriaan D. de Jongh: Drafting/revision of the manuscript for content, including medical writing for content; Major role in the acquisition of data; Study concept or design; Analysis or interpretation of data
Ruben P.A. van Eijk: Drafting/revision of the manuscript for content, including medical writing for content; Major role in the acquisition of data; Study concept or design; Analysis or interpretation of data
Leonhard A. Bakker: Drafting/revision of the manuscript for content, including medical writing for content; Analysis or interpretation of data
Tommy M. Bunte: Drafting/revision of the manuscript for content, including medical writing for content
Anita Beelen: Drafting/revision of the manuscript for content, including medical writing for content
Conny van der Meijden: Drafting/revision of the manuscript for content, including medical writing for content
Michael A. van Es: Drafting/revision of the manuscript for content, including medical writing for content
Johanna M.A. Visser-Meily: Drafting/revision of the manuscript for content, including medical writing for content
Esther T. Kruitwagen: Drafting/revision of the manuscript for content, including medical writing for content
Jan H. Veldink: Drafting/revision of the manuscript for content, including medical writing for content
Leonard H. van den Berg: Drafting/revision of the manuscript for content, including medical writing for content; Major role in the acquisition of data; Study concept or design; Analysis or interpretation of data

Figure Count:
5

Table Count:
2

Search Terms:
[177] Anterior nerve cell disease, [178] Amyotrophic lateral sclerosis

Acknowledgment:

Study Funding:
This study was funded by the Netherlands ALS Foundation (Grant No. TRICALS-Origin).

Disclosure:
The authors report no relevant disclosures.

Preprint DOI:

Received Date:
2022-12-08

Accepted Date:
2023-04-18

Handling Editor Statement:
Submitted and externally peer reviewed. The handling editor was Associate Editor Anthony Amato, MD, FAAN.

ABSTRACT (330 / 350 words)

Background and Objectives:

Current scales used in amyotrophic lateral sclerosis (ALS) attempt to summarize different functional domains or 'dimensions' into one overall score, which may not accurately characterize the individual patient's disease severity or prognosis. Use of composite scores risks declaring treatments ineffective if not all dimensions of ALS disease progression are impacted equally. We aimed to develop the ALS Impairment Multidomain Scale (AIMS) to comprehensively characterize disease progression and increase the likelihood of identifying effective treatments.

Methods:

The ALS functional rating scale (ALSFRS-R) and a preliminary questionnaire, based on literature review and patient input, were completed online by patients from the Netherlands

ALS registry at bimonthly intervals over a period of 12 months. A 2-week test-retest, factor analysis, Rasch analysis and a signal-to-noise optimization strategy were performed to create a multidomain scale. Reliability, longitudinal decline and associations with survival were evaluated. The sample size required to detect a 35% reduction in progression rate over 6 or 12 months was assessed for a clinical trial that defines the ALSFRS-R or AIMS subscales as a primary endpoint family.

Results:

The preliminary questionnaire, consisting of 110 questions, was completed by 367 patients. Three unidimensional subscales were identified and a multidomain scale was constructed with 7 bulbar, 11 motor and 5 respiratory questions. Subscales fulfilled Rasch model requirements, with excellent test-retest reliability of 0.91-0.94 and a strong relationship with survival ($p<0.001$). Compared to the ALSFRS-R, signal-to-noise ratios were higher as patients declined more uniformly per subscale. Consequently, the estimated sample size reductions achieved with the AIMS compared to the ALSFRS-R were 16.3% and 25.9% for 6- and 12-month clinical trials, respectively.

Discussion:

We developed the AIMS, consisting of unidimensional bulbar, motor and respiratory subscales that may characterize disease severity better than a total score. AIMS subscales have high test-retest reliability, are optimized to measure disease progression and are strongly related to survival time. The AIMS can be easily administered and may increase the likelihood of identifying effective treatments in ALS clinical trials.

INTRODUCTION

Amyotrophic lateral sclerosis (ALS) is a heterogeneous, multifaceted neurodegenerative disease with multiple underlying pathophysiological mechanisms and differential clinical phenotypes.[1, 2] The revised ALS functional rating scale (ALSFRS-R) is most commonly used to evaluate disease severity, monitor disease progression and serve as primary endpoint in clinical trials,[3] as it is easy to administer and strongly predictive of survival.[4, 5]

The ALSFRS-R is, however, multidimensional, meaning that multiple independent facets of ALS disease progression ('dimensions', i.e. bulbar, motor and respiratory functioning), are summarized into one total score. The fundamental problem is that patients with equal total scores may not be comparable in their disease severity or prognosis, which complicates the assessment of disease progression and treatment effects.[6-9] Alternatives have been developed, such as the ALS severity scale and Rasch-Built Overall ALS Disability Scale (ROADS),[10, 11] that similarly summarize different ALS symptoms into one total score. A total score may not, however, accurately characterize disease severity of all different ALS phenotypes.[2, 6, 9] Bulbar and spinal onset patients, for example, have different disease courses,[8] respiratory insufficiency may occur at any timepoint and many patients will never develop bulbar symptoms or weakness in all limbs.[2, 12] Additionally, treatment effects measured by a total score can become diluted, when treatments do not impact all ALS domains equally.[7]

Analyzing ALSFRS-R subscales separately may characterize disease progression more comprehensively, but does not solve inherent measurement problems, as many ALSFRS-R item options are never the most probable answer during the course of the disease.[6, 10] Moreover, the ALSFRS-R domains are ordinal instead of linearly weighted, meaning that a 1-point decline can represent either a small or large loss of functional ability depending on the question. Rasch-built scales ensure that weighting is linear and that worse answer options progressively become more probable during the course of the disease. In fact, Rasch analysis,

combined with longitudinal evaluation of candidate questions, may further improve development of a more sensitive outcome measure for ALS clinical trials.

In order to maximize the likelihood of identifying effective treatments and improve the utility of questionnaires to monitor disease progression, alternative scales are needed that account for multidimensionality, satisfy Rasch measurement standards and maximize changes over time. In this study, therefore, we aimed to develop the ALS Impairment Multidomain Scale (AIMS) to characterize disease progression comprehensively.

METHODS

Questionnaire development

A preliminary ALS disability questionnaire was created using literature review, international guidelines for ALS, clinical judgement of a panel of experts, and patient input. The literature review included existing scales and questionnaires that measure ALS function or disability,[3, 11, 13-16] and guidelines and reviews[17-20] describing ALS symptoms. The expert-panel consisted of three neurologists (MvE, JV, LvB) and two senior researchers (RvE, AB), all with extensive expertise in ALS and neuromuscular diseases. The aim was to compile a complete set of questions that covers the full range of disease progression and disability levels in ALS. The preliminary questionnaire consisted of 110 questions, each with five answer options on a Likert-type scale, similar to the Center for Neurological Study Bulbar Function Scale (CNS-BFS).[13] Subsequently, think-aloud interviews were conducted with seven patients to assess the acceptability, clarity, intelligibility and completeness of the questionnaire. After completing these interviews, questions were adjusted linguistically if patients did not fully understand them. The final questionnaire was translated into English by a professional interpreter and back-translated into Dutch for validation. This Dutch translation was compared with the original Dutch version of the AIMS and checked for inconsistencies. The

final English and Dutch versions of the AIMS can be found in eAppendices 1 and 2, respectively.

Participants

In total, 486 ALS patients, enrolled in the Netherlands ALS registry, who had previously consented to be approached for research purposes, were sent a link to the preliminary AIMS questionnaire and validated patient-reported version of the ALSFRS-R[3, 4], via e-mail, on 11 October 2019 (**Figure 1**). The population-based Netherlands ALS registry has been registering ALS patients prospectively since 2006; it has been described in detail elsewhere.[21, 22] In brief, patients diagnosed with ALS, according to the revised El Escorial or Gold Coast criteria,[23, 24] were identified via annual screening of hospital registries, specialized ALS rehabilitation clinic registries and by contacting neurologists individually. Survival time (defined as the time between enrolment and date of death or date last known to be alive) was obtained for all patients by checking the municipal register at quarterly intervals. As of 11 October 2021, which was the cut-off date for the survival analysis, 167 patients (45.5%) had died, and 2 patients (0.54%) were censored administratively with less than 6 months' follow-up time.

Study procedures

After completing the preliminary questionnaire, participants received a second link to the preliminary questionnaire to be completed within 14 days in order to evaluate test-retest reliability per question. Questions with high test-retest reliability (see below) were selected for the longitudinal phase, which required the patients to complete the questions and the validated patient-reported version of the ALSFRS-R[3, 4] every two months during 12-months' follow-up. All study data were input and stored in an online database using CASTOR Electronic Data Capture software.[25]

Reliability and Rasch analyses

Test-retest reliability was assessed by calculating the intraclass coefficient (ICC) for questions that were completed twice within 14 days; questions with an ICC less than 0.80 were removed. Exploratory factor analysis with varimax rotations was used to identify ALS domains and model fit was assessed by the root mean square error of approximation (RMSEA); an RMSEA of <0.08 was considered acceptable.[6] Questions were grouped in subscales according to the pattern of factor loadings. Principal components and factor analysis were performed to evaluate unidimensionality per subscale, which was defined as variance explained by the measured construct of >50%.[26] For the Rasch analyses, one observation per patient was randomly sampled from their longitudinal measurements to avoid dependency in the data. Rasch analyses were performed using the partial credit model,[27] allowing each question to have its own category probability curves.[28] Ideally, category probability curves should demonstrate that as disease progresses and disability increases, worse question responses sequentially become more likely. If not, question thresholds are disordered. Category probability curves were examined and any questions with disordered thresholds were removed. Differential item functioning occurs when different groups of patients with the same overall disability level answer questions significantly differently.[28] We compared questionnaire responses according to sex, age and site of symptom onset (i.e. bulbar or spinal onset) using a likelihood ratio test adjusted for multiple testing, and questions that showed significant differential item functioning were removed. Question misfit was evaluated by mean-square fit statistics and by comparing the observed proportions with 95% confidence intervals per question response with the predicted probabilities of the Rasch model; misfitting questions, demonstrating significant dependency or unmodeled noise, were removed. Correlations between questions were assessed to avoid interdependent questions.

Optimization to measure disease progression

Per subscale, longitudinal rates of decline were estimated using linear mixed effects models with a fixed effect for time and a random intercept and slope for time per patient. Average monthly rate of decline was assessed by the fixed effect of time, whereas between-patient variability was defined as the standard deviation of the random effects for time (i.e. individual progression rates). The signal-to-noise ratio was defined as the ratio between rate of decline and between-patient variability. To allow direct comparison of between-patient variability in rate of decline with the ALSFRS-R, scores were standardized by subtracting the mean and dividing by the standard deviation. Importantly, this linear transformation does not impact the ratio between the rate of decline and between-patient variability or the required sample size. Signal-to-noise ratios of the individual bulbar, motor and respiratory subscales were optimized by minimizing the sample size required to detect a given treatment effect in a clinical trial for all possible combinations of questions,[29, 30] and by selecting the combination of questions that leads to the lowest required sample size. The required sample size is a result of a combination of the rate of decline (i.e. the 'signal'), and the within- and between-patient variance components (i.e. the 'noise'). Sample size calculations were based on 80% power to detect a 35% reduction in rate of decline during 6 or 12 months' follow-up, using monthly follow-up and a two-sided alpha of 5%. Sample size calculations were performed in a subset of patients more comparable to common clinical trial populations (defined as 'trial-eligible patients'), i.e. after exclusion of patients with disease duration >36 months, age >80 years or use of non-invasive ventilation at enrolment.

The final bulbar, motor and respiratory subscales and question difficulties were reviewed by the expert panel for content validity and clinical utility to measure disability and disease progression. Empirical power of the ALSFRS-R and final AIMS subscales to detect a uniform

35% reduction in rate of decline was estimated, using an analytical strategy that evaluates treatment effects per subscale, prior to stating whether a treatment is effective, while adjusting p-values for multiple testing using the Hommel method, as described previously.[7] Empirical power of the ALSFRS-R and AIMS was estimated by resampling (n=25,000) longitudinal data of 75 patients per arm with replacement. Average rate of decline of one sampled arm was then reduced by 35% to simulate a hypothetical treatment effect, and individual ALSFRS-R and AIMS subscale scores were recalculated. In each resampled dataset, we calculated a p-value for the between-group difference in rate of decline measured by the ALSFRS-R and AIMS subscales. ALSFRS-R and AIMS subscales were defined as a primary endpoint family, i.e. a statistically significant treatment effect on any one of the subscales was considered a positive trial. Empirical power for the ALSFRS-R and AIMS was defined as the proportion of 25,000 resampled datasets with a statistically significant between-group difference in rate of decline. To make the results easier to understand, we translated empirical power to required sample size to achieve 80% power.[31]

Construct validity

Construct validity was assessed by evaluating the associations of the AIMS subscales with the ALSFRS-R and survival time. Linear mixed effects models containing the bulbar, motor and respiratory subscales as dependent variable and the corresponding ALSFRS-R subscales as fixed effects were used to evaluate associations with the ALSFRS-R. Non-linear relationships were modeled using quadratic fixed effects per ALSFRS-R subscale, and a random slope and intercept were used per patient. Bootstrapping (n=25,000) was used to estimate 95% confidence intervals. Associations of the subscales score at baseline with survival time were assessed using the Kaplan-Meier estimator and Cox regression.

Standard Protocol Approval, Registration and Patient Consent

The medical ethics committee and institutional review board of the University Medical Center Utrecht approved this study (reference 19/463) and all participants provided informed consent prior to participating.

Data availability statement

Anonymized data not published within this article will be shared on request from any qualified investigator.

RESULTS

Study population

An overview of how the questionnaire was developed is given in **Figure 1**. The preliminary questionnaire, consisting of 110 questions, and the self-reported version of the ALSFRS-R were sent to 486 ALS patients enrolled in the Netherlands ALS registry; 367 patients (75.5%) provided informed consent and completed at least one questionnaire. In total, 2,144 questionnaires were completed during 12 months' follow-up with a mean of 5.8 questionnaires and 9.3 months' follow-up time per patient. Characteristics of the study population are presented in **Table 1**. One-hundred-and-thirty-nine (37.9%) patients fulfilled the definition of trial eligibility based on a disease duration of less than 36 months, age younger than 80 years and no use of non-invasive ventilation at enrolment. Trial-eligible patients were slightly younger, and had a shorter disease duration, better ALSFRS-R score at inclusion, but a faster decline. The average rate of decline in ALSFRS-R total score was 0.63 (95% CI 0.56 to 0.71) points per month for all patients and 1.02 (95% CI 0.88 to 1.17) points per month for the trial-eligible patients.

The median time to complete the ALSFRS-R and preliminary 120-item questionnaire was 18 minutes (interquartile range, 12 – 30). **eFigure 1** shows that all disease stages were represented at all time points as ALSFRS-R total scores ranged from 0 to 48.

Rasch analyses

The number of patients who completed a second preliminary questionnaire within 14 days for the test-retest analysis was 146. Of the 110 questions, 67 (61%) had an ICC of 0.80 or higher, thereby surpassing the selection threshold and were subsequently collected during 12-months' follow-up. Principal components and factor analysis found that three unidimensional domains were sufficient to explain the majority (55.2%) of variance in the data, while this was 49.2% and 59.2% for two and four domains, respectively. The pattern of factor loadings suggested that questions represented three separate domains: a bulbar, a motor, and a respiratory domain.

Category probability curves that describe the probability of each question response per bulbar, motor and respiratory disability level were examined and the three intermediate answer options were collapsed to resolve disordered thresholds, resulting in a total of three response options per question. Five questions ('dietary changes due to swallowing difficulties', 'drooling', 'use of walking aid', 'need help getting out of bed', 'use of air-stacking') were removed, as collapsing response categories did not solve the problem of disordered thresholds. When comparing questionnaire responses between men and women, no significant differential item functioning was observed. Three questions were removed due to differential item functioning; one question ('repeating myself to be understood') demonstrated differential functioning due to age and two questions ('people that understand me tell other people what I said' and 'I am aware of my speech disorder') demonstrated differential functioning on the basis of site of symptom onset. Lastly, model fit per question was assessed and six questions

('in the morning I see saliva on my pillow', 'walking is exhausting', 'standing up', 'changing leg position', 'use of analgesics' and 'use of a wheelchair') were removed due to question misfit.

Optimization to measure disease progression

Bulbar, motor and respiratory subscales were each optimized to measure ALS disease progression. In total, there were 53 remaining questions that assessed bulbar (n=14), motor (n=29) or respiratory (n=10) functioning. **Figure 2** shows the relationship between the total number of questions per subscale and the required sample size. Initially, adding more questions increases information (i.e. increases average rate of decline, reduces between-patient variability, reduces within-patient variability or a combination of the three), thus reducing the sample size required to detect a given treatment effect in a clinical trial. However, at some point an optimum is reached, where adding more questions does not lead to an increase in information but increases 'noise', and hence to an increase in the sample size required to detect treatment effects. The final combination of questions that resulted in the lowest required sample size (or within 5% of the minimum) consisted of 7 bulbar, 11 motor and 5 respiratory questions. Compared to the ALSFRS-R subscales, the AIMS subscales reduced the 12-month sample size by 23.9%, 27.6% and 53.6% **(Table 2)**. Next, we estimated the sample size reductions for a clinical trial that defines the ALSFRS-R or AIMS subscales as a primary endpoint family, i.e. by evaluating treatment effects univariately per bulbar, motor and respiratory subscale while adjusting p-values for multiple testing, prior to determining whether a treatment is effective overall. In this case, a statistically significant treatment effect on any one of the subscales was considered a positive trial. Compared to the ALSFRS-R subscales, estimated sample size reductions were 16.3% and 25.9%, respectively, for a 6-month and 12-month clinical trial. Results were similar in sensitivity analyses that included patients less comparable to common trial populations (i.e. including patients with

disease duration >36 months, being older than 80 years or use of non-invasive ventilation at baseline) (**eTable 1**). Question difficulties are presented in **Figure 3,** showing that worse question options sequentially become more probable as disability per subscale increases. Importantly, the AIMS targeted a broader range of ALS disability levels than the ALSFRS-R, with larger question location disability ranges. Question locations, expressed as logits on a Rasch disability scale, ranged from -0.69 to 1.97, -1.76 to 1.53 and -1.10 to 1.65, for bulbar, motor and respiratory subscales, respectively, while the corresponding ALSFRS-R subscale question location ranges were 0.18 to 0.94, -0.18 to 1.37 and -0.78 to 0.42.

Reliability and construct validity

Test-retest reliability for bulbar, motor and respiratory subscale scores was high with ICCs of 0.94 (95%CI 0.91 to 0.95), 0.94 (95%CI 0.92 to 0.96) and 0.91 (95%CI 0.88 to 0.94), respectively. Finally, to assess construct validity, we evaluated the AIMS subscale score associations with corresponding ALSFRS-R subscales and survival time after enrolment (**Figure 4**). Correlations with respective ALSFRS-R bulbar, motor and respiratory subscores was 0.87 (95%CI 0.85 to 0.90), 0.93 (95%CI 0.92 to 0.94) and 0.79 (95%CI 0.75 to 0.82). Compared with the ALSFRS-R, ceiling and floor effects of the AIMS appeared to be smaller. For example, a patient with an ALSFRS-R bulbar score of 0 has, on average, an AIMS bulbar score of 4 (**Figure 4A**). AIMS subscales were associated with overall survival, lower scores resulting in lower survival probabilities after enrolment, all subscales $p<0.001$. Or, using Cox regression, hazard ratios for bulbar, motor and respiratory subscales were 0.90 (95%CI 0.86 to 0.94), 0.94 (95%CI 0.92 to 0.97) and 0.84 (95%CI 0.79 to 0.90), respectively, all $p < 0.001$. Or in other words, a 1-point increase in bulbar, motor or respiratory score was associated with a 10%, 6% and 16% reduction in risk of death, respectively.

DISCUSSION

In this study, we developed the ALS Impairment Multidomain Scale, consisting of unidimensional subscales with 7 bulbar, 11 motor and 5 respiratory questions, that have high test-retest reliability, fulfill Rasch requirements and are strongly related to survival time. We optimized subscales by selecting questions that optimize longitudinal rate of decline, while reducing between-patient variability. We have thus developed an easily administered outcome measure for use in ALS clinical trials and in the clinic, that may be more sensitive with a broader measurement window than the ALSFRS-R for monitoring disease progression and detecting treatment effects.

Importantly, by developing a multidomain scale, we aimed to address the multidimensional nature of ALS symptoms. Regulatory agencies guidance suggests that for some disorders, multiple endpoints may be required for full characterization of the disease.[32] Multidimensionality is a feature of many neurological diseases, such as multiple sclerosis,[33] spinal muscular atrophy,[34] Parkinson's disease,[35] stroke[36] and Alzheimer's disease.[37] Current scales in ALS, such as the ALSFRS-R,[3] ALS severity scale[11] or ROADS,[10] summarize a range of ALS symptoms in one composite score, which may not accurately characterize a multidimensional disease like ALS. Due to a heterogeneous clinical presentation and different sub-phenotypes, patients with the same ALSFRS-R total score are not necessarily comparable with regard to disease severity, progression rate or prognosis.[6, 8, 38] Moreover, using a composite total score as primary endpoint in clinical trials can disguise important treatment clues. Treatments may not impact all ALS domains equally; as a result, treatment effects measured by composite endpoints may become diluted. For example, in the Nuedexta trial,[39] treatment only improved bulbar functioning (p=0.003); this treatment effect may have been missed if the ALSFRS-R total score (p=0.25) was defined as primary endpoint. Obviously, in this study, the beneficial effect of Nuedexta on bulbar functioning was hypothesized

beforehand, but the effect is often not known a priori. Similarly, in the edaravone and sodium phenylbutyrate-taurursodiol trials, the treatment effect on the ALSFRS-R total score seemed to be primarily driven by the motor subscale rather than bulbar and respiratory subscales.[40, 41] Likewise, studies that focus on non-drug interventions, such as optimizing multidisciplinary care via physical therapy, exercise programs or speech therapy may also benefit from a more comprehensive assessment of the effects of the intervention.[42] The multidomain AIMS may, therefore, better characterize ALS disease progression and treatment effects, and subsequently facilitate disease monitoring in both the clinic and in trials.

There are several analytical strategies for analyzing multidomain scales that avoid the pitfalls of composite endpoints, while controlling the false positive rate (i.e. type I error).[43] A relatively straightforward method is to evaluate treatment effects per subscale, prior to stating whether a treatment is effective overall, while adjusting p-values for multiple testing.[7] Depending on the investigator's preference, AIMS subscales can a priori be defined as a primary endpoint family,[32] i.e. a treatment effect on any one of the subscales is considered clinically relevant and may be indicative of treatment effectiveness (**eTable 2** provides a worked example). Another option is to rank the importance of each AIMS subscale. This can be done on a group-level (e.g. bulbar is always more important than motor function),[44] or using individual patient or physician preferences.[45] The treatment effect can subsequently be summarized as the probability of obtaining a more favorable outcome when treated compared to when receiving placebo. **Figure 5** illustrates how this type of analysis could be presented for the AIMS. Nevertheless, as presented in **eTable 3**, more complex analytical strategies exist, such as a prespecified testing hierarchy or using multivariate models.[46, 47] Multivariate models simultaneously model multiple longitudinal outcomes, allowing calculation of one p-value for the overall treatment effect, which may be a more powerful strategy than performing multiple independent statistical tests. Multivariate models are flexible, as longitudinal

outcomes can be added (e.g. vital capacity, muscle strength or biomarkers) and, importantly, could be adjusted for mortality.[46, 48]

Rasch analysis was used to systematically reduce the number of questions in the preliminary questionnaire, to ensure that the AIMS is linearly weighted and that patients can be compared using subscale scores as only one concept is being measured (i.e. bulbar, motor or respiratory disability). To further optimize the multidomain scale, we analyzed longitudinal decline during 12 months' follow-up and selected the combination of questions that results in a high average rate of decline (i.e. 'the signal') with minimal variability (i.e. 'the noise').[29, 30] Consequently, compared to the corresponding ALSFRS-R subscales, the signal-to-noise ratio was improved (**Table 2**), thus increasing the precision of the AIMS in monitoring disease progression. Reducing variability in trial endpoints is important as it increases precision in estimating treatment effects and reduces the sample size required to detect a given treatment effect.

This study has several limitations that need to be considered. First, cognitive impairment was not evaluated in our study. It is, therefore, not clear if and to what extent cognitive impairment influenced results for the ALSFRS-R and AIMS. Second, examination of category probability curves indicated collapsing intermediate question responses was necessary to avoid disordered thresholds, making prospective validation of the newly phrased questions necessary. Third, in the current study, the AIMS was patient-reported, making it easy to incorporate as a remote survey in clinical trials. We have, however, further refined the AIMS by providing guidance per question on when to score 0, 1 or 2. We hypothesize that providing such guidance in combination with adequate training of research personnel may reduce variability in responses within and between patients. However, whether providing guidance and training for scoring indeed reduces variability should be investigated in future studies. Lastly, we found strong associations of the AIMS subscales with survival time.

However, for a questionnaire to be regarded as a true surrogate endpoint for survival time, it is important that a treatment effect on survival time is reflected by the surrogate endpoint and vice versa. Future prospectively designed studies, that, for example, use the joint modelling framework,[46, 48] are, therefore, important in establishing the relationship between treatment effects on the AIMS, ALSFRS-R and survival time.

In conclusion, we have developed the AIMS, consisting of unidimensional bulbar, motor and respiratory subscales that may characterize disease severity better than a total score. AIMS subscales have high test-retest reliability, are optimized to measure disease progression and are strongly related to survival time. The AIMS can be easily administered and may increase the likelihood of identifying effective treatments in ALS clinical trials.

WNL-2023-000313_sup ---http://links.lww.com/WNL/C898

eAppendix AIMS Dutch ---http://links.lww.com/WNL/C899

eAppendix AIMS English --- http://links.lww.com/WNL/C900

REFERENCES

1. van Es MA, Hardiman O, Chio A, Al-Chalabi A, Pasterkamp RJ, Veldink JH, van den Berg LH. Amyotrophic lateral sclerosis. Lancet 2017;390:2084-2098.

2. Swinnen B, Robberecht W. The phenotypic variability of amyotrophic lateral sclerosis. Nat Rev Neurol 2014;10:661-670.

3. Cedarbaum JM, Stambler N, Malta E, Fuller C, Hilt D, Thurmond B, Nakanishi A. The ALSFRS-R: A revised ALS functional rating scale that incorporates assessments of respiratory function. BDNF ALS study group (phase III). J Neurol Sci 1999;169:13-21.

4. Bakker LA, Schroder CD, Tan HHG, et al. Development and assessment of the inter-rater and intra-rater reproducibility of a self-administration version of the ALSFRS-R. J Neurol Neurosurg Psychiatry 2020;91:75-81.

5. Kimura F, Fujimura C, Ishida S, et al. Progression rate of ALSFRS-R at time of diagnosis predicts survival time in ALS. Neurology 2006;66:265-267.

6. Franchignoni F, Mora G, Giordano A, Volanti P, Chio A. Evidence of multidimensionality in the ALSFRS-R scale: A critical appraisal on its measurement properties using rasch analysis. J Neurol Neurosurg Psychiatry 2013;84:1340-1345.

7. van Eijk RPA, de Jongh AD, Nikolakopoulos S, McDermott CJ, Eijkemans MJC, Roes KCB, van den Berg LH. An old friend who has overstayed their welcome: The ALSFRS-R total score as primary endpoint for ALS clinical trials. Amyotroph Lateral Scler Frontotemporal Degener 2021:1-8.

8. Rooney J, Burke T, Vajda A, Heverin M, Hardiman O. What does the ALSFRS-R really measure? A longitudinal and survival analysis of functional dimension subscores in amyotrophic lateral sclerosis. J Neurol Neurosurg Psychiatry 2017;88:381-385.

9. Bakker LA, Schröder CD, van Es MA, Westers P, Visser-Meily JMA, van den Berg LH. Assessment of the factorial validity and reliability of the ALSFRS-R: A revision of its measurement model. J Neurol 2017;264:1413-1420.

10. Fournier CN, Bedlack R, Quinn C, et al. Development and validation of the rasch-built overall amyotrophic lateral sclerosis disability scale (ROADS). JAMA Neurol 2020;77:480-488.

11. Hillel AD, Miller RM, Yorkston K, McDonald E, Norris FH, Konikow N. Amyotrophic lateral sclerosis severity scale. Neuroepidemiology 1989;8:142-150.

12. Fujimura-Kiyono C, Kimura F, Ishida S, Nakajima H, Hosokawa T, Sugino M, Hanafusa T. Onset and spreading patterns of lower motor neuron involvements predict survival in sporadic amyotrophic lateral sclerosis. J Neurol Neurosurg Psychiatry 2011;82:1244-1249.

13. Smith RA, Macklin EA, Myers KJ, et al. Assessment of bulbar function in amyotrophic lateral sclerosis: Validation of a self-report scale (center for neurologic study bulbar function scale). Eur J Neurol 2018;25:907-e66.

14. Helleman J, Kruitwagen-van Reenen ET, Bakers J, et al. Using patient-reported symptoms of dyspnea for screening reduced respiratory function in patients with motor neuron diseases. J Neurol 2020;267:3310-3318.

15. Appel V, Stewart SS, Smith G, Appel SH. A rating scale for amyotrophic lateral sclerosis: Description and preliminary experience. Ann Neurol 1987;22:328-333.

16. Vogt S, Petri S, Dengler R, Heinze HJ, Vielhaber S. Dyspnea in amyotrophic lateral sclerosis: Rasch-based development and validation of a patient-reported outcome (DALS-15). J Pain Symptom Manage 2018;56:736-745.e2.

17. Miller RG, Jackson CE, Kasarskis EJ, et al. Practice parameter update: The care of the patient with amyotrophic lateral sclerosis: Drug, nutritional, and respiratory therapies (an evidence-based review): Report of the quality standards subcommittee of the american academy of neurology. Neurology 2009;73:1218-1226.

18. Hobson EV, McDermott CJ. Supportive and symptomatic management of amyotrophic lateral sclerosis. Nat Rev Neurol 2016;12:526-538.

19. National Institute for Health and Care Excellence. NICE guideline motor neuron disease: Assessment and management [online]. Available at: https://www.nice.org.uk/guidance/ng42. Accessed 06/01, 2019.

20. Fang T, Jozsa F, Al-Chalabi A. Nonmotor symptoms in amyotrophic lateral sclerosis: A systematic review. Int Rev Neurobiol 2017;134:1409-1441.

21. de Jongh AD, van Eijk RPA, Peters SM, et al. Incidence, prevalence, and geographical clustering of motor neuron disease in the netherlands. Neurology 2021;96:e1227-e1236.

22. Huisman MH, de Jong SW, van Doormaal PT, et al. Population based epidemiology of amyotrophic lateral sclerosis using capture-recapture methodology. J Neurol Neurosurg Psychiatry 2011;82:1165-1170.

23. Brooks BR, Miller RG, Swash M, Munsat TL, World Federation of Neurology Research Group on Motor Neuron Diseases. El escorial revisited: Revised criteria for the diagnosis of

amyotrophic lateral sclerosis. Amyotroph Lateral Scler Other Motor Neuron Disord 2000;1:293-299.

24. Shefner JM, Al-Chalabi A, Baker MR, et al. A proposal for new diagnostic criteria for ALS. Clin Neurophysiol 2020;131:1975-1978.

26. Linacre JM. Rasch analysis of rank-ordered data. J Appl Meas 2006;7:129-139.

27. Masters GN. A rasch model for partial credit scoring. Psychometrika 1982;47:149-174.

28. Vanhoutte EK, Hermans MC, Faber CG, Gorson KC, Merkies IS, Thonnard JL, PeriNomS Study Group. Rasch-ionale for neurologists. J Peripher Nerv Syst 2015;20:260-268.

29. de Jongh AD, van den Berg LH, van Eijk RPA. Reconsidering the revised amyotrophic lateral sclerosis functional rating scale for ALS clinical trials. J Neurol Neurosurg Psychiatry 2020.

30. Ard MC, Edland SD. Power calculations for clinical trials in alzheimer's disease. J Alzheimers Dis 2011;26 Suppl 3:369-377.

31. Healy BC, Schoenfeld D. Comparison of analysis approaches for phase III clinical trials in amyotrophic lateral sclerosis. Muscle Nerve 2012;46:506-511.

32. U.S. Department of Health and Human Services Food and Drug Administration Center for Drug Evaluation and Research (CDER). Multiple endpoints in clinical trials guidance for industry [online]. Available at: https://www.fda.gov/media/102657/download. Accessed 4/14, 2022.

33. Hobart J, Lamping D, Fitzpatrick R, Riazi A, Thompson A. The multiple sclerosis impact scale (MSIS-29): A new patient-based outcome measure. Brain 2001;124:962-973.

34. Mazzone E, De Sanctis R, Fanelli L, et al. Hammersmith functional motor scale and motor function measure-20 in non ambulant SMA patients. Neuromuscul Disord 2014;24:347-352.

35. Martinez-Martin P, Gil-Nagel A, Gracia LM, Gomez JB, Martinez-Sarries J, Bermejo F. Unified parkinson's disease rating scale characteristics and structure. the cooperative multicentric group. Mov Disord 1994;9:76-83.

36. Rethnam V, Bernhardt J, Johns H, et al. Look closer: The multidimensional patterns of post-stroke burden behind the modified rankin scale. Int J Stroke 2021;16:420-428.

37. Robert P, Ferris S, Gauthier S, Ihl R, Winblad B, Tennigkeit F. Review of alzheimer's disease scales: Is there a need for a new multi-domain scale for therapy evaluation in medical practice? Alzheimers Res Ther 2010;2:24.

38. Al-Chalabi A, Chio A, Merrill C, Oster G, Bornheimer R, Agnese W, Apple S. Clinical staging in amyotrophic lateral sclerosis: Analysis of edaravone study 19. J Neurol Neurosurg Psychiatry 2021;92:165-171.

39. Smith R, Pioro E, Myers K, et al. Enhanced bulbar function in amyotrophic lateral sclerosis: The nuedexta treatment trial. Neurotherapeutics 2017;14:762-772.

40. Paganoni S, Macklin EA, Hendrix S, et al. Trial of sodium phenylbutyrate-taurursodiol for amyotrophic lateral sclerosis. N Engl J Med 2020;383:919-930.

41. Clinical review report: Edaravone (radicava): (Mitsubishi tanabe pharma corporation). 2019.

42. van Groenestijn AC, Schröder CD, van Eijk RPA, et al. Aerobic exercise therapy in ambulatory patients with ALS: A randomized controlled trial. Neurorehabil Neural Repair 2019;33:153-164.

43. Ristl R, Urach S, Rosenkranz G, Posch M. Methods for the analysis of multiple endpoints in small populations: A review. J Biopharm Stat 2019;29:1-29.

44. Evans SR, Rubin D, Follmann D, et al. Desirability of outcome ranking (DOOR) and response adjusted for duration of antibiotic risk (RADAR). Clin Infect Dis 2015;61:800-806.

45. van Eijk RPA, van den Berg LH, Lu Y. Composite endpoint for ALS clinical trials based on patient preference: Patient-ranked order of function (PROOF). J Neurol Neurosurg Psychiatry 2021.

46. Rizopoulos D, Ghosh P. A bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event. Stat Med 2011;30:1366-1380.

47. Pinheiro J, Bates D, DebRoy S, Sarkar D, Team RC. Nlme: Linear and nonlinear mixed effects models. 2013.

48. van Eijk RP, Eijkemans MJ, Rizopoulos D, van den Berg LH, Nikolakopoulos S. Comparing methods to combine functional loss and mortality in clinical trials for amyotrophic lateral sclerosis. Clin Epidemiol 2018;10:333-341.

Figure 1. Flowchart of study design and questionnaire development steps

Figure 1 legend. The flowchart shows the study and questionnaire development steps used in this study. A preliminary questionnaire with 110 questions was sent to ALS patients enrolled in the Netherlands ALS registry. Test-retest reliability, longitudinal follow-up, Rasch analysis and optimization to measure disease progression using longitudinal data were performed to create the final unidimensional bulbar, motor and respiratory subscales of the AIMS.

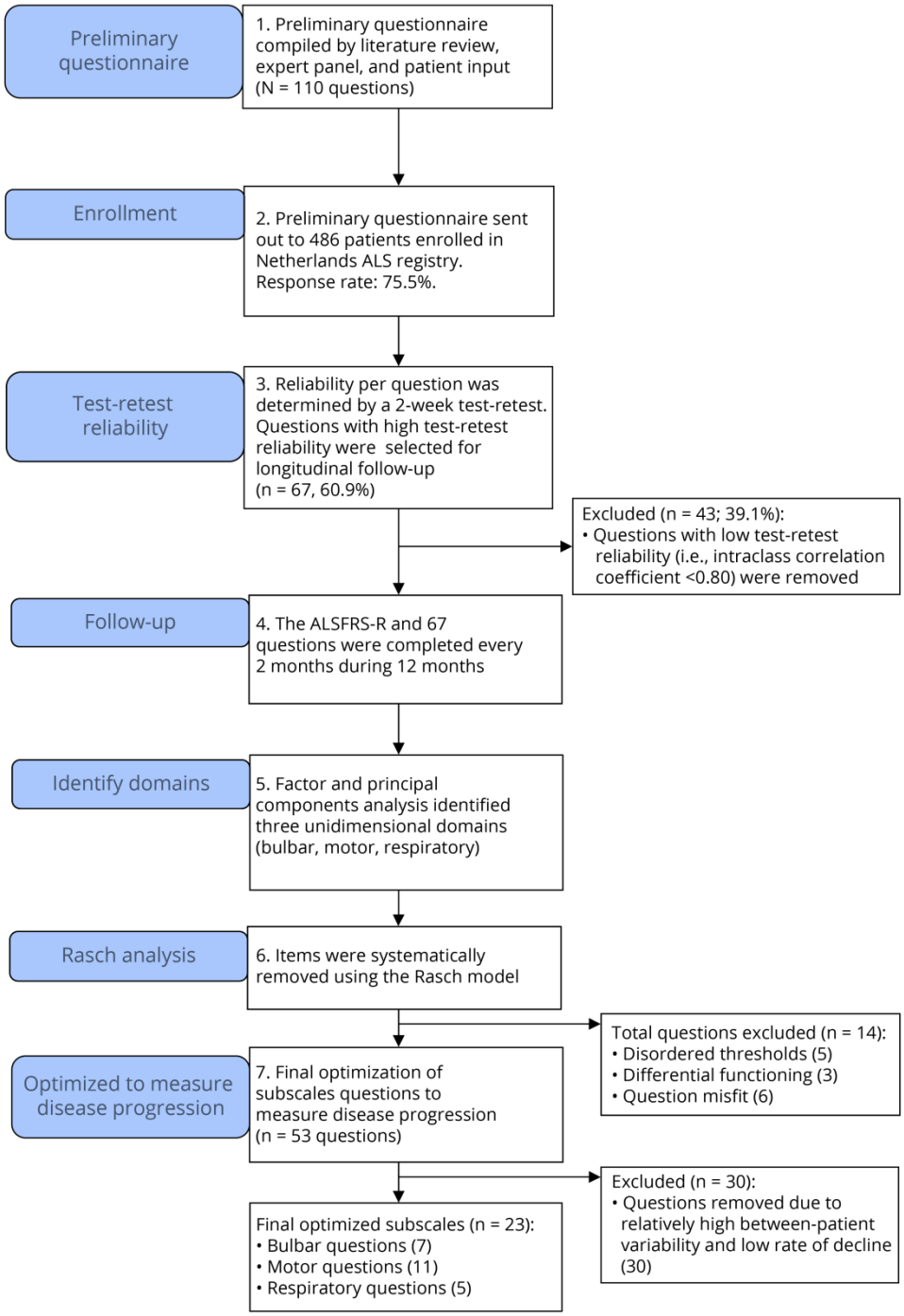


Table 1. Characteristics of study participants at baseline

	All patients *N = 367*	**Trial-eligible patients** *N = 139*
Age, years	65 (11)	62 (10)
Male sex	251 (68%)	90 (65%)
Spinal onset	298 (81%)	103 (75%)
Disease duration, months	37 (21 – 75)	19 (14 – 26)
Diagnostic delay, months	15 (9 – 33)	10 (6 – 15)
ALSFRS-R total score	33 (26 – 40)	38 (32 – 42)
ΔFRS, points per month	0.33 (0.18 – 0.61)	0.51 (0.30 – 0.86)
Riluzole use	245 (67%)	107 (77%)
Respiratory support	83 (23%)	0 (0%)
Gastrostomy	36 (10%)	8 (5.8%)
Frontotemporal dementia	20 (5.3%)	7 (5.5%)

Data are n (%), mean (standard deviation) or median (interquartile range). Presence of frontotemporal dementia was determined at time of diagnosis as cognitive status was unknown at enrolment. The trial-eligible patients subgroup was defined as patients with disease duration <36 months, no use of non-invasive ventilation at baseline and being younger than 80 years. ALSFRS-R total score was calculated as the sum of items 1-12.
ΔFRS was calculated as (48 – ALSFRS-R total score at enrolment) / disease duration at enrolment in months.

Figure 2. Number of questions versus required sample size per subscale

Figure 2 legend. The required sample size for a 12-month clinical trial to detect a 35% reduction in progression rate with 80% power was used as an estimate of the sensitivity of candidate subscales to measure disease progression. For each number of questions, the combination of questions that resulted in the lowest required sample size is plotted. Solid rectangles indicate candidate subscales that result in the lowest (+5%) sample size. Initially, as questions are added, information increases, thus reducing the required sample size. However, as more similar questions are subsequently added, between-patient variability in rate of decline increases, thus inflating the sample size required to detect treatment effects.

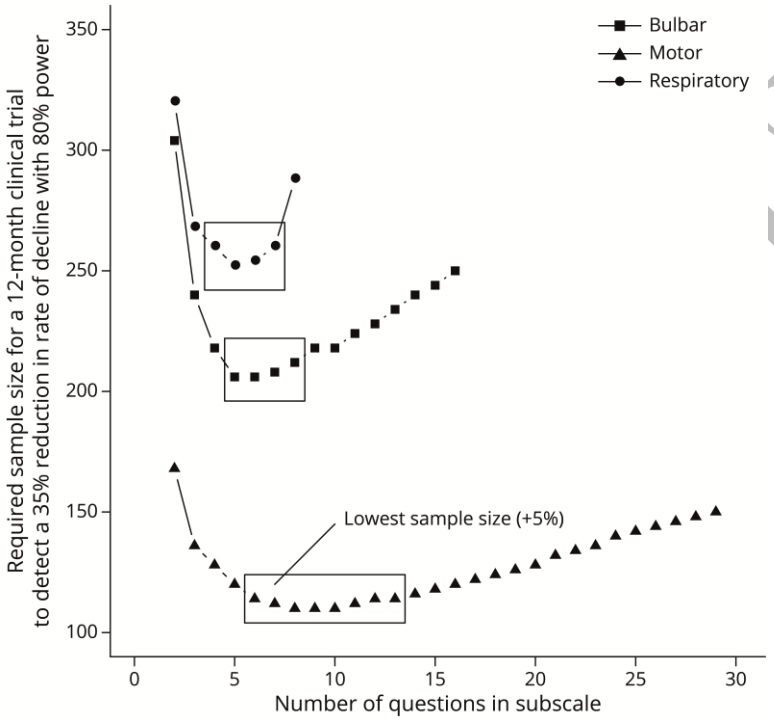


Figure 3. AIMS question locations and thresholds

Figure 3 legend. The figure shows the ALS Impairment Multidomain Scale (AIMS), question difficulty order and targeting of each question. The most difficult question (e.g. chewing and swallowing) is on the far right, while the easiest question (e.g. respiratory insufficiency) is on the far left. The x-axis represents the logit measure for a patient's bulbar, motor or respiratory disability. Worse question options sequentially become more probable as disability increases.

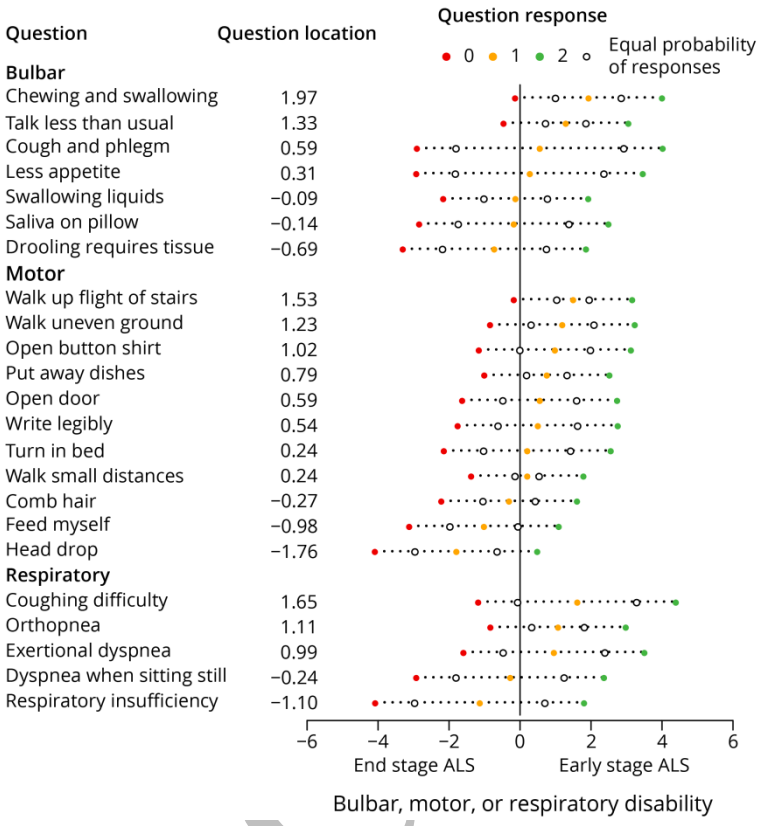


Table 2. Longitudinal decline and required sample size per ALSFRS-R and AIMS subscale in the trial-eligible patients

	No. of questions	**Baseline score** *(intercept)*	**Rate of decline** *(points per month)*	**Between-patient variability** *(σ_b)*	**Within-patient variability** *(σ_w)*	**Signal-to-noise ratio**	**Required sample size 6-month study**	**Required sample size 12-month study**
ALSFRS-R bulbar subscale	3	0.20 (9.9)	0.061 (0.21)	0.062 (0.21)	0.23 (0.78)	0.99	392	284
AIMS bulbar subscale	7	0.20 (10.9)	0.062 (0.20)	0.052 (0.16)	0.30 (0.93)	1.18	392 (-0.0%)	216 (-23.9%)
ALSFRS-R motor subscale	6	0.26 (16.0)	0.084 (0.55)	0.071 (0.47)	0.19 (1.23)	1.17	234	196
AIMS motor subscale	11	0.27 (14.0)	0.084 (0.50)	0.060 (0.36)	0.23 (1.34)	1.40	198 (-15.4%)	142 (-27.6%)
ALSFRS-R respiratory subscale	3	0.39 (11.6)	0.10 (0.24)	0.14 (0.34)	0.51 (1.20)	0.70	758	560
AIMS respiratory subscale	5	0.36 (8.8)	0.095 (0.17)	0.089 (0.16)	0.46 (0.83)	1.06	440 (-42.0%)	260 (-53.6%)

We performed the analysis after exclusion of patients less likely to enroll in trials (i.e. disease duration >36 months, age >80 years and use of non-invasive ventilation at baseline). Linear mixed effects models were used to analyze longitudinal trajectories per subscale. Scores were standardized by subtracting the mean and dividing by the standard deviation to allow direct comparison between subscales; rates of decline and variability in the original units are given between brackets. Sample size calculations were based on 80% power to detect a 35% reduction in rate of decline in a 6- or 12-month clinical trial using monthly follow-up.

Figure 4. Association of AIMS subscales with the ALSFRS-R and survival time

Figure 4 legend. Bulbar, motor and respiratory AIMS scores were strongly associated with the corresponding ALSFRS-R subscales (A-C) and survival time (D-F). There was a dose-response association between baseline scores and overall survival, lower scores resulting in lower survival probabilities after enrolment.

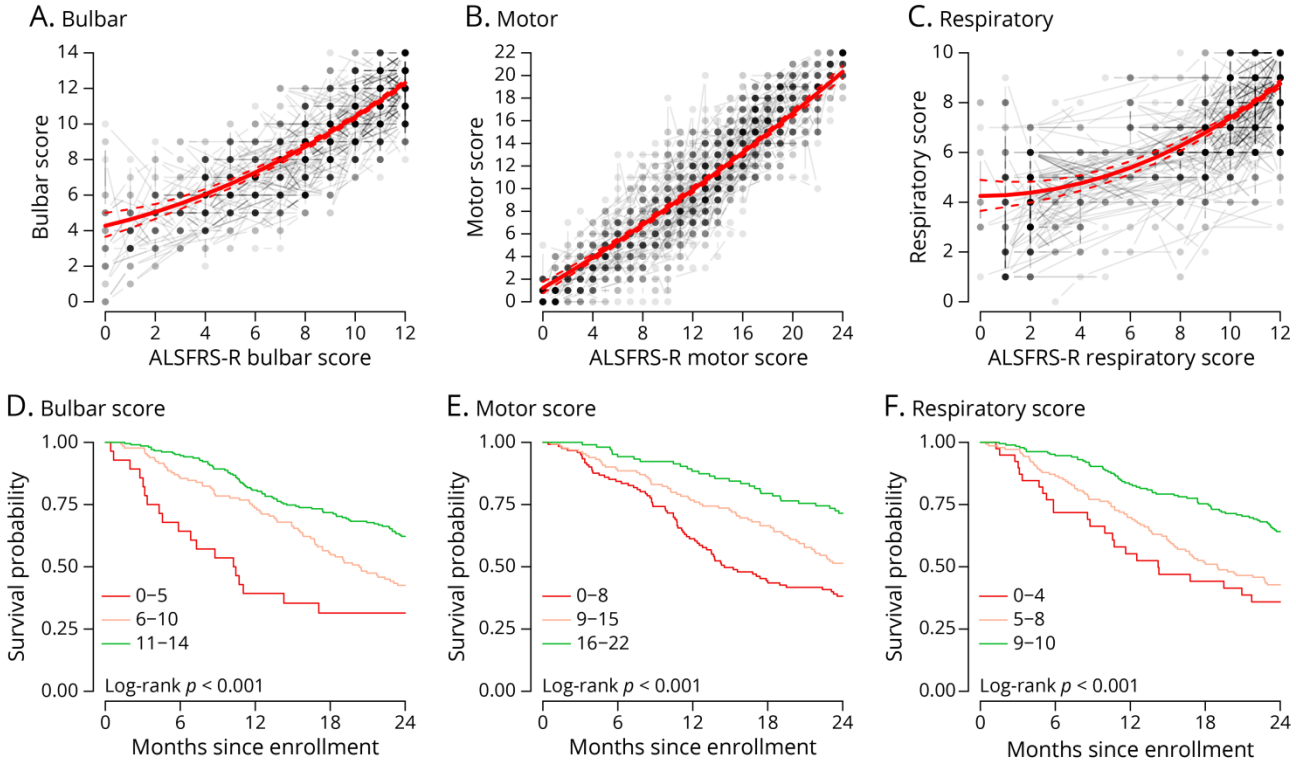
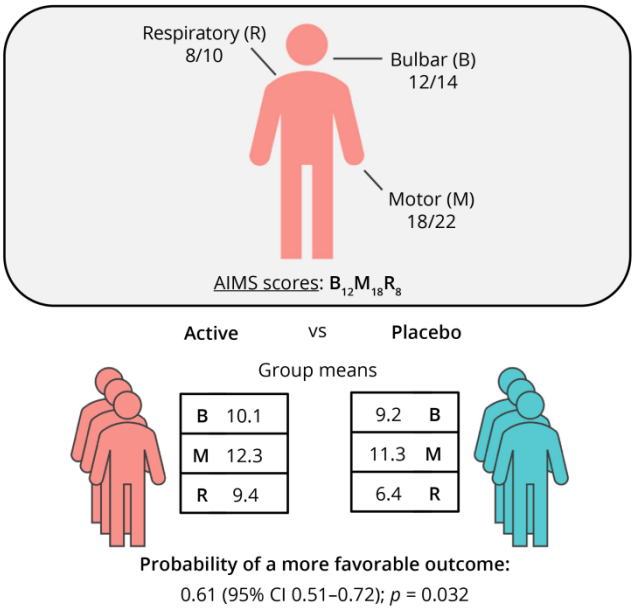


Figure 5. Example reporting of the AIMS in ALS clinical studies

Figure 5 legend. In this example study, the AIMS subscales were used as primary endpoint. By weighing each of the domains, for example, according to patient, physician or prognostic preferences, one can obtain one overall effect size with one overall p-value. A similar approach is used by the Combined Assessment of Function and Survival, thereby prioritizing one outcome (survival time) over the other (ALSFRS-R). This approach results in one overall effect size, in this case the probability or odds of having a better outcome when receiving the new therapy compared to receiving placebo. Other strategies for analyzing multiple domains are presented in eTable 3.



Development of a Rasch-Built Amyotrophic Lateral Sclerosis Impairment Multidomain Scale to Measure Disease Progression in ALS

Adriaan D. de Jongh, Ruben P.A. van Eijk, Leonhard A. Bakker, et al.

Neurology published online June 13, 2023
DOI 10.1212/WNL.0000000000207483