Original research

Composite endpoint for ALS clinical trials based on patient preference: Patient-Ranked Order of Function (PROOF)

Ruben P A van Eijk ![ORCID],[1,2] L H van den Berg,[2] Ying Lu[1]

[1]Department of Biomedical Data Science and Centre for Innovative Study Design, Stanford University, Stanford, California, USA
[2]Department of Neurology, UMC Utrecht Brain Centre, Utrecht, The Netherlands

Correspondence to
Dr Ruben P A van Eijk, Department of Neurology, UMC Utrecht Brain Centre, Utrecht, The Netherlands; R.P.A.vanEijk-2@umcutrecht.nl

Check for updates

ABSTRACT

Background Patients with amyotrophic lateral sclerosis (ALS) show considerable variation in symptoms. Treatments targeting an overall improvement in symptomatology may not address what the majority of patients consider to be most important. Here, we propose a composite endpoint for ALS clinical trials that weighs the improvement in symptoms compared with what the patient population actually wants.

Methods An online questionnaire was sent out to a population-based registry in The Netherlands. Patients with ALS were asked to score functional domains with a validated self-reported questionnaire, and rank the order of importance of each domain. This information was used to estimate variability in patient preferences and to develop the Patient-Ranked Order of Function (PROOF) endpoint.

Results There was extensive variability in patient preferences among the 433 responders. The majority of the patients (62.1%) preferred to prioritise certain symptoms over others when evaluating treatments. The PROOF endpoint was established by comparing each patient in the treatment arm to each patient in the placebo arm, based on their preferred order of functional domains. PROOF averages all pairwise comparisons, and reflects the probability that a patient receiving treatment has a better outcome on domains that are most important to them, compared with a patient receiving placebo. By means of simulation we illustrate how incorporating patient preference may upgrade or downgrade trial results.

Conclusions The PROOF endpoint provides a balanced patient-focused analysis of the improvement in function and may help to refine the risk–benefit assessment of new treatments for ALS.

INTRODUCTION

Patients with amyotrophic lateral sclerosis (ALS) are affected in multiple domains limiting their bulbar, arm, leg and/or respiratory function.[1 2] The revised ALS functional rating scale (ALSFRS-R) is the most commonly used method for assessing these domains.[3] Its total score has become the recommended coprimary efficacy endpoint, next to survival, for clinical trials in ALS.[4 5] Similar scoring methods have been proposed, such as the Rasch Overall ALS Disability Scale,[6] or the Appel ALS score.[7] What these scores have in common is the intention to summarise a variety of symptoms 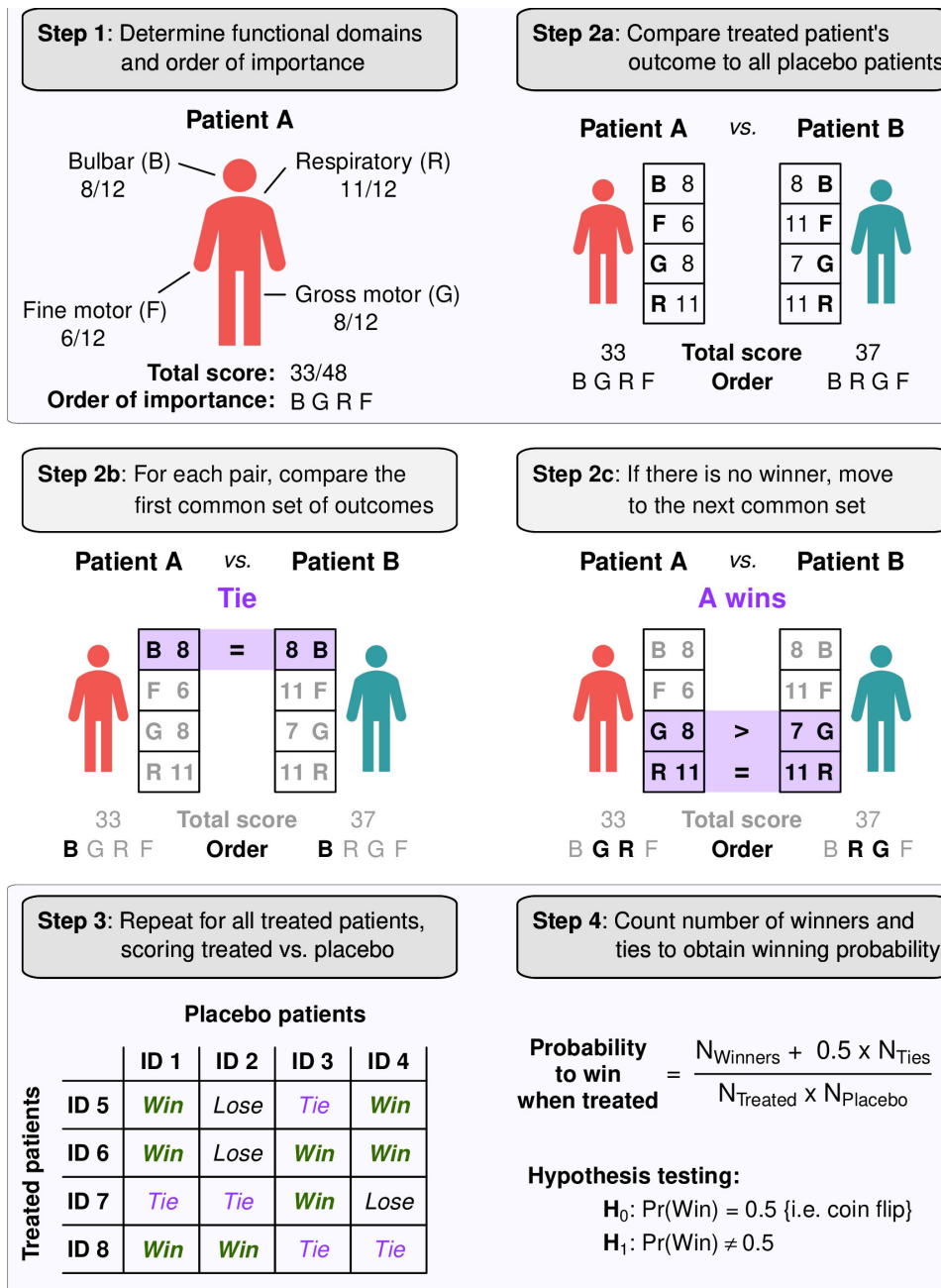into total scores that reflect overall function. Beyond the discussion about whether these scores fulfil exact psychometric or clinimetric requirements,[6 8 9] a more fundamental challenge arises if patients do not consider the symptoms to be of equal importance.

For example, not all patients with a bulbar site of symptom onset will develop symptoms in the legs, even in the final stages of the disease.[10] In fact, most patients who die from ALS have an ALSFRS-R total score that is higher than zero,[11] indicating that some symptoms may never develop during the course of the disease. From a patient perspective, therefore, it is not evident that all symptoms assessed by scores such as the ALSFRS-R are of equal importance, and patients may prioritise some symptoms over others. At a group level, this could affect how valuable new treatments are for patients: it could be contested whether a treatment that addresses symptoms of little concern to patients, despite a beneficial effect in ALSFRS-R total score, is really beneficial and whether the benefits outweigh the potential risks.

In order to make an accurate assessment of the value of new treatments for ALS, there is a need to consider the totality of benefits and harms, including what is of primary concern to patients.[12] Yet, current clinical trial endpoints may not reflect what the patient considers to be most important and might overvalue or undervalue the benefit of new treatments. In this study, therefore, we propose a new composite endpoint for randomised controlled clinical trials in ALS based on the patient preference for functional domains, thereby weighing the treatment effect compared with what the patient population actually wants.

METHODS

Participants

Patients in The Netherlands with a diagnosis within the spectrum of motor neuron diseases, at all stages of disease and irrespective of cognitive impairment, were approached by email. In The Netherlands, patients diagnosed with either possible, probable (laboratory supported) or definite ALS according to the revised El Escorial criteria,[13] or diagnosed with progressive muscular atrophy (PMA) or primary lateral sclerosis (PLS) are registered centrally at The Netherlands ALS Centre.[14] Vital status is updated at quarterly intervals by checking the online municipal population register. In May 2021, we approached all patients in the registry who were alive on 1 May 2021 (N=1243) and had provided prior consent to be re-contacted for medical research purposes (789

Figure 1 Schematic illustration to obtain winning probability when treated. The Patient-Ranked Order of Function endpoint compares each treated patient to each placebo patient. Based on the preferred order of functional domains, one determines whether the treated patient has a better, worse or equal outcome compared with the placebo patient. The comparisons are summarised into a winning probability, reflecting the probability that a random patient receiving treatment has a more desirable outcome than a random patient receiving placebo. Examples and solutions of individual patient comparisons are available online (https://tricals.shinyapps.io/PROOF/)

out of 1243, 63.5%). A valid e-mail address was available for 668 of the 789 patients (84.7%).

Study procedures

An online questionnaire was constructed using a cloud-based clinical data management platform (Castor EDC, V.2021.2, https://www.castoredc.com). The questionnaire consisted of a validated self-reported version of the ALSFRS-R.[15] This consists of twelve items that can be clustered into four domains: (1) bulbar; items 1–3, (2) fine motor; items 4–6, (3) gross motor; items 7–9 and (4) respiratory functioning; items 10–12.[16 17] Domain scores range from 0 to 12, with higher scores reflecting

better function. The online questionnaire was supplemented with two additional questions: 'Which domain bothers you the most?' and 'Imagine you will receive a treatment that delays disease progression; delay of which domain is the most important to you?'. The latter could be answered by ranking the four domains in their order of importance using ranks 1 (most important) to 4 (least important), or by indicating that there is no preference (ie, all domains are of equal importance). Automated data validations were programmed to minimise missing data and ensure data quality. In case of non-response, a one-time reminder was sent after 10 days, and the database was locked after 30 days.

Table 1 Patient characteristics of the enrolled population

Characteristic	All patients (N=433)	Stratified by MND-subtype		
		ALS (N=271)	PMA (N=74)	PLS (N=88)
Age at study enrolment, years	65 (10)	64 (10)	67 (9)	66 (10)
Sex, male	295 (68%)	186 (69%)	57 (77%)	52 (60%)
Symptom duration,* months	51 (78)	36 (40)	79 (78)	143 (125)
Diagnostic delay,* months	17 (25)	12 (16)	26 (47)	34 (49)
Site of symptom onset, bulbar	82 (19%)	65 (24%)	2 (3%)	15 (17%)
ALSFRS-R total score	32 (10)	31 (10)	33 (10)	35 (8)
ΔFRS, points per month	−0.37 (.38)	−0.48 (0.41)	−0.26 (0.28)	−0.13 (0.19)
Riluzole use, yes	271 (63%)	214 (79%)	46 (62%)	11 (13%)
Gastrostomy, yes	67 (16%)	58 (21%)	5 (7%)	4 (5%)
Respiratory support, yes	87 (20%)	62 (23%)	18 (24%)	7 (8%)
Presence of frontotemporal dementia†	9 (2%)	7 (3%)	0 (0%)	2 (2%)

Data are expressed as mean (SD) or n (%).
*Data are expressed as median (IQR).
†Determined at diagnosis, cognitive status unknown at time of questionnaire.
ALS, amyotrophic lateral sclerosis; ALSFRS-R, ALS functional rating scale; ΔFRS, ALSFRS-R − 48/symptom duration; MND, motor neuron disease; PLS, primary lateral sclerosis; PMA, progressive muscular atrophy.

The original questionnaire is available online (https://tricals.shinyapps.io/PROOF/).

Patient-Ranked Order of Function endpoint

Consider a randomised controlled clinical trial comparing an experimental treatment with a placebo arm. The Patient-Ranked Order of Function (PROOF) endpoint was defined such that each treated patient is compared with each placebo patient. Based on the preferred order of functional domains, one determines whether the treated patient has a better, worse or equal outcome compared with the placebo patient (figure 1).[12] For each pair of patients (ie, one patient from the treated arm and one patient from the placebo arm), the comparison starts with the set of domains that the pair commonly ranked as most important. As illustrated in figure 1, both patient A and patient B find the bulbar domain most important and the bulbar domain score is used for the comparison. A patient can be scored as winner, loser or equal to its comparator. In case of a tie, the comparison is moved to the second set of domains that the pair commonly ranked as most important, else the comparison is completed and the next pair of patients is assessed. Here, patient A and patient B have equal bulbar scores, and the comparison is moved to the next set of domains (respiratory and gross motor). In the case of a comparison with multiple domains, a patient wins if at least one domain scores higher and none of the other domains scores lower. A patient loses if at least one domain scores lower, and none of the other domains scores higher. Likewise, there is a tie when at least one domain scores lower and at least one domain scores higher, or if all domains are equal. In case of a tie, the comparison continues with the next common set of domains, or if all domains have been compared, the total scores of patients are compared; this completes the comparison. If both patients indicated that they had no preference for any of the domains, only total scores are used for the comparison and the domain scores are disregarded. If only one patient has a preferred order, while the other has no preference, domains are compared in the order as indicated by the patient with a preference. An online calculator was developed to illustrate the winner for any pair of patients given their preferences and domain scores (https://tricals.shinyapps.io/PROOF/).

Statistical analysis and interpretation of PROOF

Ultimately, all treated patients are compared with all placebo patients. This results in a matrix with the number of rows equal to the number of treated patients, and the number of columns equal to the number of placebo patients (figure 1, step 3). Subsequently, one counts the total number of wins plus half the number of ties, resulting in the U statistic of the experimental group and its significance can be determined using the non-parametric Mann-Whitney U test. To make interpretation easier, we divided the U statistic by the number of comparisons (ie, the number of cells in the matrix), resulting in a winning probability. The winning probability simply reflects the probability that a random patient receiving treatment has a *more desirable* outcome than a random patient receiving placebo. If there is no benefit or harm from treatment, the winning probability is 0.5 or equal to flipping a coin. If the treatment harms patients, the winning probability is less than 0.5. The online calculator (https://tricals.shinyapps.io/PROOF/) contains background information about its derivation.

Comparison with the ALSFRS-R total score

To illustrate the differences between PROOF and the ALSFRS-R total score in our cohort, we extended the comparison by comparing each patient in our dataset to all other patients following the PROOF algorithm. For each patient, we counted his or her points (1 for each win, 0.5 for each tie, 0 for each loss) and ranked patients accordingly. The patient with rank 1 had the lowest score, reflecting the 'overall loser' of the cohort, whereas the patient ranked highest reflected the 'overall winner'. As alternative, we repeated the process but now scoring patients only on the basis of their ALSFRS-R total score. Simple scatter plots and boxplots were used to illustrate differences between the two endpoints.

In addition, a simulation exercise was conducted to illustrate the effect of the PROOF algorithm on clinical trials, by assuming that our questionnaire responses reflected postinterventional trial scores in a virtual placebo arm. We resampled with replacement patients from the questionnaire responses and randomly allocated patients to a virtual placebo or treatment group. If the patient was allocated to treatment, we added a fixed constant to his or her domain scores, reflecting a hypothetical treatment effect. Subsequently, we calculated the winning probability
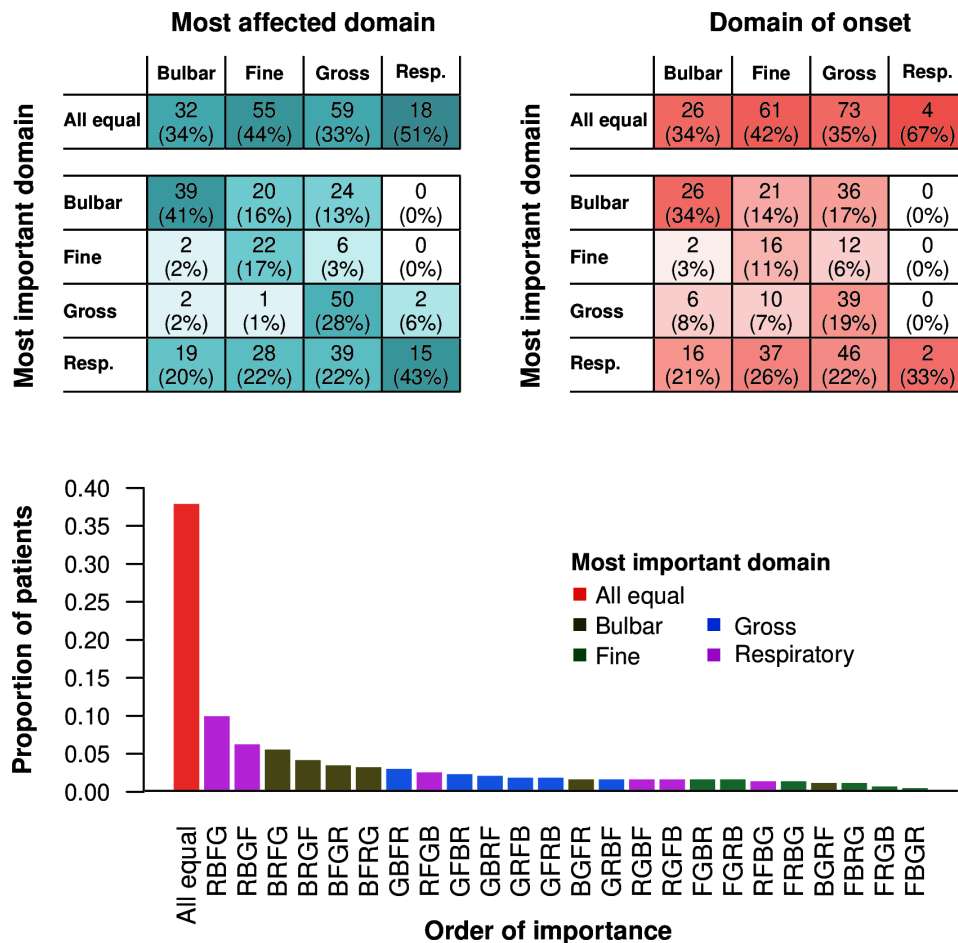
Figure 2 Distribution of patient preferences for functional domains in amyotrophic lateral sclerosis. The tables reflect the relationship between the most important domain reported by the patient and their most affected domain or domain of symptom onset. The numbers reflect the number of patients; percentages are calculated for each column. To illustrate: of all patients who indicated that their bulbar domain was most affected, 41% considered the bulbar domain to be the most important, followed by no preference (ie, all equal, 34%). The bar chart below reflects the prevalence of each order, coloured by the domain that was reported first (eg, 'RBFG' translates to Respiratory (1)—Bulbar (2)—Fine (3)—Gross (4), with the respiratory domain as most important and gross motor as least important).

according to the PROOF algorithm and the ALSFRS-R total score. The process was repeated 10 000 times to estimate empirical power, and repeated for various hypothetical treatment effects. The number of replications was determined based on the SE of a proportion, targeting a 95% accuracy of ±1%.

RESULTS
In total, 668 patients were approached, of whom 500 (74.9%) responded to the initial invitation. Of the 500 responses, 15 patients declined to participate, 32 remained indecisive and 20 patients started, but did not complete, the questionnaire. This resulted in a total sample size of 433 complete questionnaire responses, originating from 271 (63%) patients with ALS, 74 (17%) patients with PMA and 88 (20%) patients with PLS; the patient characteristics are presented in table 1. Overall, the patient population reflects a typical prevalence cohort, with a wide variety in duration of symptoms, disease severity and clinical stages (eg, ALSFRS-R total scores ranging from 0 to 48).

Variability in patient preferences
The majority of the patients (269 out of 433, 62.1%) indicated they preferred prioritising certain domains over others when evaluating treatments. This percentage was 45.8% (38 out of

83) for patients in an early disease stage (ie, King's clinical stage 1)[18] and 66.0% (231 out of 350) for patients in a later disease stage (ie, King's clinical stage >1, p=0.001). Of the patients with a preference, the most important domain reported by the patient was respiratory (37.5%), followed by bulbar (30.9%), gross motor (20.4%) and fine motor (11.2%). Figure 2 illustrates the relationship between the most important domain and the most affected domain at the time of the questionnaire, or the domain of first symptom onset. For patients with a preference, there is a weak to moderate association between the domain that is most important to them and the domain that is most affected (Cohen's Kappa 0.323, p<0.001), or was first affected (Cohen's Kappa 0.121, p<0.001), as reflected by the higher proportions in the table's diagonal. Overall, a wide variation was found in the order of importance of domains reported by patients (*lower bar chart*). The orders which appeared most frequently started with either respiratory or bulbar functioning. Results are similar when patients with PMA or PLS are excluded (*not shown*).

PROOF and the ALSFRS-R total score
In figure 3A, we represent the association between patients ranked according to the PROOF algorithm and the ALSFRS-R total score in our cohort. Despite the strong relationship
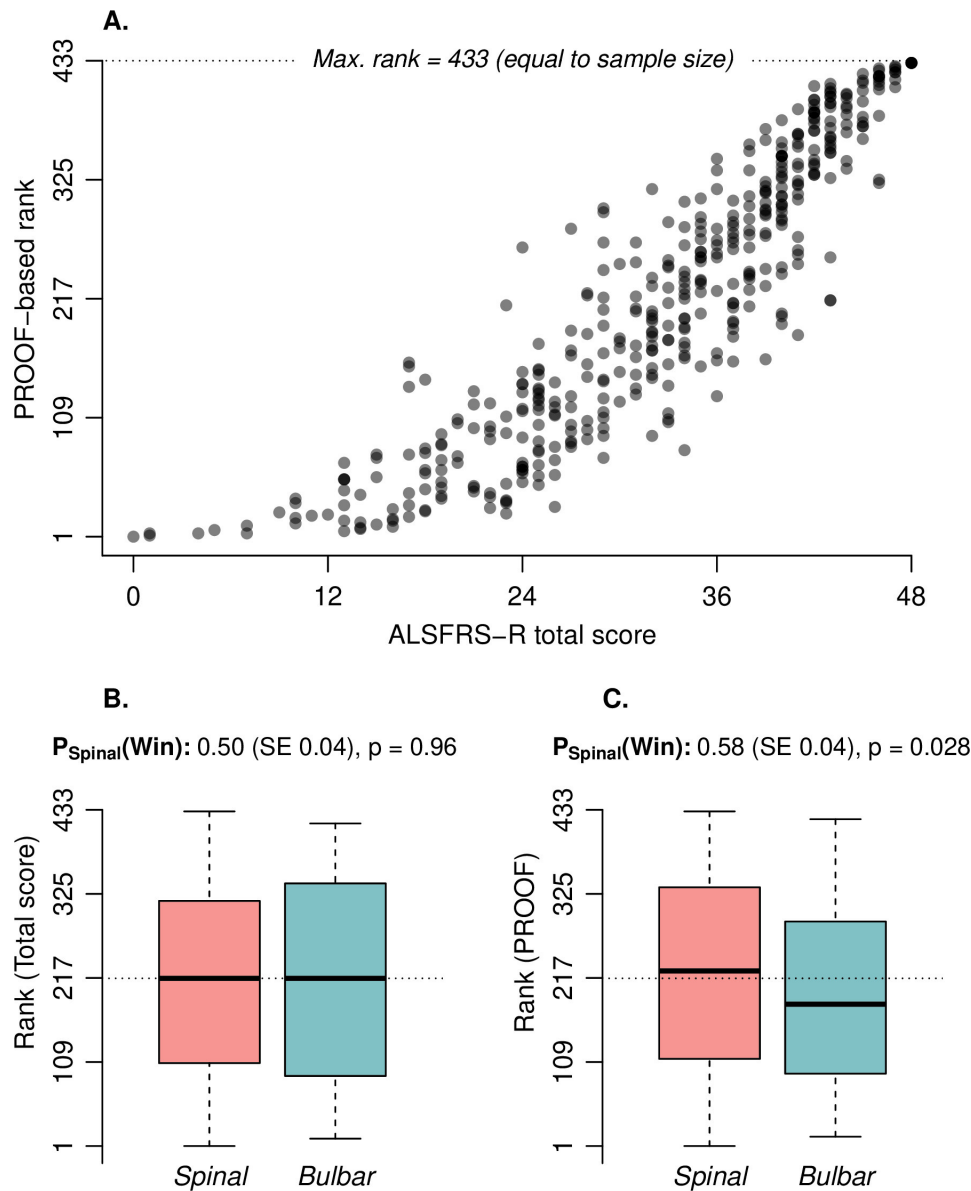
Figure 3 Relationship between PROOF and ALSFRS-R total score. Each patient in our dataset was compared with all other patients following the algorithm illustrated in figure 1 (*step 2*). For each patient, we counted his or her points (1 for winner, 0.5 for tie, 0 for loser); patients with the most points were, subsequently, ranked highest (ranks range from 1 to 433). Patients with the highest ALSFRS-R total score also have the highest PROOF-based rank, and vice versa (A). Patients were ranked by either ALSFRS-R total score alone (B) or by PROOF (C). To illustrate: we stratified the ranks according to site of symptom onset, showing that PROOF may result in group differences even if total scores are comparable. Scores were compared using the Mann-Whitney U test; the winning probability was calculated as $U_{Spinal}/N_{Spinal}N_{Bulbar}$. ALSFRS-R, amyotrophic lateral sclerosis functional rating scale; PROOF, Patient-Ranked Order of Function.

between the two endpoints (Spearman rho 0.929, p<0.001), PROOF creates contrasts between patients who have identical ALSFRS-R total scores. For example, there were 12 patients with an ALSFRS-R total score of 46 out of 48, for whom PROOF-based ranks ranged from 322 to 426. In this case, the patient ranked as 322nd scored 10 out of 12 points on bulbar, whereas the patient ranked 426th scored 10 out of 12 points on gross motor functioning. As the bulbar domain is ranked to be of higher importance than the gross motor domain, bulbar function loss is considered worse than loss of gross motor function, resulting in differentiating scores on the PROOF endpoint, despite the identical ALSFRS-R total scores. Similarly, at a group-level, this can lead to contrasts even when groups appear to be identical when the ALSFRS-R total score is used (figure 3B,C). Note that

if there is no difference between groups, the winning probability is 0.50 or equal to flipping a coin. Here, the probability that a spinal patient will have higher scores on domains that are most important to them, compared with a bulbar patient, is 0.58 (95% CI 0.51 to 0.65, p=0.028), whereas the probability that a spinal patient will have a higher ALSFRS-R total score compared with a bulbar patient is 0.50 (95% CI 0.43 to 0.57, p=0.96).

PROOF as clinical trial endpoint
Finally, in table 2 we provide, by means of simulation, a comparison between the PROOF endpoint and the ALSFRS-R total score for hypothetical outcomes that may be observed during a clinical trial. If there is no benefit of treatment, the probability

Table 2 Comparison of proof and the ALSFRS-R total score in a hypothetical randomised controlled clinical trial

Hypothetical treatment effect on subdomain scores at end of trial

Description	Bulbar	Fine	Gross	Resp.	ALSFRS-R total score		PROOF	
					Winning probability	Power	Winning probability	Power
There is no difference between treatment arms in any subdomain or in total scores (null hypothesis)	+0	+0	+0	+0	0.500	0.050	0.500	0.050
There is a 4-point difference between treatment arms in total score, where all subdomains increase by one point when treated	+1	+1	+1	+1	0.611	0.479	0.630	0.618
There is a 4-point difference between treatment arms in total score, where the bulbar and respiratory subdomains increase by two points when treated	+2	+0	+0	+2	0.610	0.476	0.652	0.758
There is a 4-point difference between treatment arms in total score, where the fine and gross motor subdomains increase by two points when treated	+0	+2	+2	+0	0.611	0.489	0.579	0.278

ALSFRS-R total scores were compared using the Mann-Whitney U-test; the winning probability was calculated as $U_{treated}/N_{treated} N_{placebo}$. Results are based on 10 000 simulations by resampling 50 'placebo' and 50 'experimental' patients from the ALS population described in table 1.
ALS, amyotrophic lateral sclerosis; ALSFRS-R, ALS functional rating scale.

that a treated patient will have a higher ALSFRS-R total score, or higher scores in the domains that are most important to them, compared with a placebo patient, is 0.50. If treatment leads to a 4-point gain in the ALSFRS-R total score, this increases the probability of a treated patient having a higher ALSFRS-R total score, compared with a placebo patient, to 0.61, irrespective of how treatment affects the underlying domains. In contrast, the PROOF endpoint 'upgrades' or 'downgrades' the winning probability, depending on how treatment affects the domains that are most important to patients. If a treatment affects the domains that are more important to patients (*row 3*), this leads to a higher winning probability and increased statistical power compared with the ALSFRS-R total score. If a treatment is beneficial, but affects domains that are of lesser importance to patients (*row 4*), PROOF reduces the contrast between treatment arms and reduces the probability of considering such treatment as being beneficial.

DISCUSSION

In this study, we show that the majority of the patients with ALS prefer to prioritise certain domains over others when evaluating treatments. The extensive variability in patient preferences indicates that simply summarising an overall improvement in function may not be sufficient to accurately reflect the true value of a treatment for patients. Here, we propose a new composite endpoint for ALS clinical trials to quantify the patient-level benefit of a drug. The PROOF endpoint weighs the observed functional gain compared with what the patient population actually wants. As a consequence, treatments that improve the domains considered very important by patients receive higher scores than treatments affecting lesser important domains. Thus, the PROOF endpoint provides a balanced patient-focused analysis of the improvement in function and refines the risk–benefit assessment of new treatments.

Regulatory approval and market authorisation of new therapeutic interventions are based on a complete assessment of the benefits and risks that may be introduced. Increasingly, the patient's perspective and preferences are an important source of information to better inform decision-makers regarding the general unmet need, the potential value of new treatment strategies and relevant clinical trial outcomes.[19–21] An important missing link, however, is a metric that weighs the clinical outcomes compared with what is most important to patients, and which combines the preferences of patients in the overall assessment of the drug's benefit.[22] In this study, we have proposed a simple method to obtain a single numerical value that summarises the observed improvement in functional status together with the preferences of each individual patient in the study. The PROOF endpoint thereby allows regulators to make a more refined trade-off between the actual performance of a therapeutic intervention during clinical development and the potential harms.

For drug developers, however, defining the PROOF endpoint as primary objective could pose a risk if the primary interest lies in improving the ALSFRS-R total score. When patients consider the drug's domain-specific improvements to be of lesser importance, this reduces the treatment's winning probability and the study results could lose significance. On the other hand, when treatment benefits domains of higher importance, the PROOF endpoint increases the winning probability and improves statistical power. The PROOF endpoint, therefore, is a patient-focused analysis of the treatment effect and is only partially driven by the actual improvement in ALSFRS-R total score. Nevertheless, irrespective of the primary

interest, the PROOF endpoint allows drug developers to obtain more insight into how to redesign or improve their drugs to better address the needs of the patient population for whom they are intended and can be used either as a coprimary or secondary endpoint.

The PROOF endpoint, like the Combined Assessment of Function and Survival (CAFS),[23] is a non-parametric statistic that ranks patients based on pairwise comparisons. The PROOF and CAFS statistics can be conceptualised as the total number of winners in a treatment group or, when comparing groups, the difference in the number of winners between two treatment arms. The scores of these statistics depend, therefore, on the number of patients in the study, which prevents investigators from directly comparing the magnitude of the treatment effect across clinical trials. In this study, and as suggested previously,[12] we have improved the interpretation of the statistic by dividing the absolute score by the number of comparisons. The resulting winning probability has a more straightforward interpretation and allows for a direct comparison of treatment effects between studies. Nevertheless, the winning probability remains a relatively abstract metric that may be difficult to conceptualise or communicate to a patient. As with any composite endpoint, additional analyses of the domain-specific components and preferences are recommended in order to fully disentangle the treatment benefits.

Given the moderate associations between patient preference and their most affected domain, a strength of the PROOF endpoint is that it addresses the multidimensional aspect of ALS.[3 9] The individual PROOF ranks may provide a more refined evaluation of patient-level functioning compared with using an overall summary of function in which each domain is weighted equally (as illustrated in figure 3). This does not only benefit clinical trials, but may also enhance epidemiological or non-interventional research efforts. Moreover, the PROOF algorithm can be easily extended to other endpoints and is not limited to the ALSFRS-R domains or a particular measurement scale. For example, one could opt to create a composite of multiple primary and key secondary endpoints including muscle strength, lung function, cognition and quality of life (each measured in their own units), choose to use an alternative questionnaire instead of the ALSFRS-R, or use the rate of change rather than the actual scores. An important consideration, however, is to combine endpoints that are meaningful and relatable to the patient; asking a patient about their preferences for biological markers, such as changes in neurofilaments or neurophysiological parameters, may not be realistic. In addition, one can integrate a minimally clinically important difference in the PROOF algorithm, where a patient is only scored as winner when the difference with the comparator exceeds a certain threshold.[12]

In practice, the PROOF endpoint adds minimally to the study burden: on the day of randomisation, one additional question would suffice to register the patient preferences. The subsequent randomisation ensures that the patient preferences are independent of the treatment allocation. For non-randomised settings, for example, a comparison against historical data, it is important to match not only according to disease characteristics, but also on the basis of patient preferences collected prior to treatment initiation in order to prevent bias. At the end of the trial, some informative missing data due to death is to be expected.[24–26] One solution is to define an additional step in the PROOF-algorithm similar to the CAFS[23]: a patient who dies would automatically lose from a patient who survives. Alternatively, outcome data may be imputed for the final visit by using, for example, a joint model that accounts for death.[27] Finally, sample size calculation could be done conservatively by using the estimated sample size for the ALSFRS-R total score. Otherwise, a simulation-based approach could be employed; the patient-level dataset of this study is available at https://tricals.shinyapps.io/PROOF/ to guide input parameters for simulation studies.

Of note, the data collection in our study was self-reported and conducted remotely. As a relevant proportion of patients develop some degree of cognitive impairment,[28 29] this could have affected the patient's ability to provide their preferences. Moreover, the caregiver may have played a major role in the completion of the questionnaire. These limitations could have impacted the answers provided by the patient and affected the reported preferences. It would be of value to collect additional information on cognition and caregiver assistance in future patient-reported studies and decentralised clinical trials. For in-clinic trial settings, however, these limitations may be of less significance as patients with cognitive impairment are often excluded from participating, and data collection is often performed by trained healthcare professionals.

In conclusion, in this study we propose a new composite endpoint to determine the patient-level value of treatments for ALS. By combining what is of primary concern to patients with the observed clinical benefit in function, the PROOF endpoint provides a balanced patient-focused analysis of the risk–benefit assessment and can be an important step towards more patient-centric clinical trials in ALS.

ORCID iD
Ruben P A van Eijk http://orcid.org/0000-0002-7132-5967

REFERENCES

1 Al-Chalabi A, Hardiman O, Kiernan MC, et al. Amyotrophic lateral sclerosis: moving towards a new classification system. *Lancet Neurol* 2016;15:1182–94.

2 van Es MA, Hardiman O, Chio A, et al. Amyotrophic lateral sclerosis. *Lancet* 2017;390:2084–98.

3 van Eijk RPA, de Jongh AD, Nikolakopoulos S, et al. 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;22:300–7.

4 European Medicines Agency. Guideline on clinical investigation of medicinal products for the treatment of amyotrophic lateral sclerosis, 2016. Available: https://www.ema.europa.eu/documents/scientific-guideline/guideline-clinical-investigation-medicinal-products-treatment-amyotrophic-lateral-sclerosis_en.pdf [Accessed 04 May 2021].

5 Food drug administration center for drugs evaluation research. guidance for industry: amyotrophic lateral sclerosis: developing drugs for treatment, 2019. Available: https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM596718.pdf [Accessed 04 May 2021].

6 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;77:480–8.

7 Appel V, Stewart SS, Smith G, et al. A rating scale for amyotrophic lateral sclerosis: description and preliminary experience. *Ann Neurol* 1987;22:328–33.

8 Franchignoni F, Mandrioli J, Giordano A, *et al*. A further Rasch study confirms that ALSFRS-R does not conform to fundamental measurement requirements. *Amyotroph Lateral Scler Frontotemporal Degener* 2015;16:331–7.

9 Franchignoni F, Mora G, Giordano A, *et al*. Evidence of multidimensionality in the ALSFRS-R scale: a critical appraisal on its measurement properties using Rasch analysis. *J Neurol Neurosurg Psychiatry* 2013;84:1340–5.

10 Fujimura-Kiyono C, Kimura F, Ishida S, *et al*. Onset and spreading patterns of lower motor neuron involvements predict survival in sporadic amyotrophic lateral sclerosis. *J Neurol Neurosurg Psychiatry* 2011;82:1244–9.

11 Swinnen B, Robberecht W. The phenotypic variability of amyotrophic lateral sclerosis. *Nat Rev Neurol* 2014;10:661–70.

12 Lu Y, Zhao Q, Zou J. A composite endpoint approach to evaluate treatment benefits based on patient preferences. *Under review* 2021.

13 Brooks BR, Miller RG, Swash M. World Federation of Neurology Research group on motor neuron D. El Escorial revisited: revised criteria for the diagnosis of amyotrophic lateral sclerosis. *Amyotroph Lateral Scler Other Motor Neuron Disord* 2000;1:293–9.

14 Huisman MHB, de Jong SW, van Doormaal PTC, *et al*. Population based epidemiology of amyotrophic lateral sclerosis using capture-recapture methodology. *J Neurol Neurosurg Psychiatry* 2011;82:1165–70.

15 Bakker LA, Schröder 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.

16 Cedarbaum JM, Stambler N, Malta E, *et al*. 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.

17 Thakore NJ, Lapin BR, Kinzy TG, *et al*. Deconstructing progression of amyotrophic lateral sclerosis in stages: a Markov modeling approach. *Amyotroph Lateral Scler Frontotemporal Degener* 2018;19:483–94.

18 Balendra R, Jones A, Jivraj N, *et al*. Estimating clinical stage of amyotrophic lateral sclerosis from the ALS functional rating scale. *Amyotroph Lateral Scler Frontotemporal Degener* 2014;15:279–84.

19 Jimenez-Moreno AC, van Overbeeke E, Pinto CA, *et al*. Patient preferences in rare diseases: a qualitative study in neuromuscular disorders to inform a quantitative preference study. *Patient* 2021;14:601–12.

20 Food Drug Administration Center for Drugs Evaluation Research. FDA Patient-Focused Drug Development Guidance Series for Enhancing the Incorporation of the Patient's Voice in Medical Product Development and Regulatory Decision Making, 2020. Available: https://www.fda.gov/drugs/development-approval-process-drugs/fda-patient-focused-drug-development-guidance-series-enhancing-incorporation-patients-voice-medical [Accessed 30 August 2021].

21 European Medicines Agency. Ich reflection paper on proposed ICH guideline work to advance patient focused drug development, 2020. Available: https://www.ema.europa.eu/en/documents/scientific-guideline/ich-reflection-paper-proposed-ich-guideline-work-advance-patient-focused-drug-development_en.pdf [Accessed 30 Aug 2021].

22 Janssens R, Huys I, van Overbeeke E, *et al*. Opportunities and challenges for the inclusion of patient preferences in the medical product life cycle: a systematic review. *BMC Med Inform Decis Mak* 2019;19:189.

23 Berry JD, Miller R, Moore DH, *et al*. The combined assessment of function and survival (CAFS): a new endpoint for ALS clinical trials. *Amyotroph Lateral Scler Frontotemporal Degener* 2013;14:162–8.

24 van Eijk RPA, Nikolakopoulos S, Roes KCB, *et al*. Critical design considerations for time-to-event endpoints in amyotrophic lateral sclerosis clinical trials. *J Neurol Neurosurg Psychiatry* 2019;90:1331–7.

25 Pétavy F, Guizzaro L, Antunes Dos Reis I, *et al*. Beyond "Intent-to-treat" and "Per protocol": Improving assessment of treatment effects in clinical trials through the specification of an estimand. *Br J Clin Pharmacol* 2020;86:1235–9.

26 Oude Rengerink K, Mitroiu M, Teerenstra S, *et al*. Rethinking the intention-to-treat principle: one size does not fit all. *J Clin Epidemiol* 2020;125:198–200.

27 van Eijk RPA, Rooney J, Hardiman O, *et al*. Two heads are better than one: benefits of joint models for ALS trials. *J Neurol Neurosurg Psychiatry* 2019;90:1071–2.

28 Beeldman E, Govaarts R, de Visser M, *et al*. Progression of cognitive and behavioural impairment in early amyotrophic lateral sclerosis. *J Neurol Neurosurg Psychiatry* 2020;91:779–80.

29 Chiò A, Moglia C, Canosa A, *et al*. Cognitive impairment across ALS clinical stages in a population-based cohort. *Neurology* 2019;93:e984–94.