This is a repository copy of *Innovating clinical trials for amyotrophic lateral sclerosis : challenging the established order*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/180068/

Version: Published Version

# Innovating Clinical Trials for Amyotrophic Lateral Sclerosis

## Challenging the Established Order

Ruben P.A. van Eijk, MD, PhD, Stavros Nikolakopoulos, PhD, Kit C.B. Roes, PhD, Lindsay Kendall, MSc, Steve S. Han, MD, PhD, Arseniy Lavrov, MD, PhD, Noam Epstein, MD, Tessa Kliest, MSc, Adriaan D. de Jongh, MD, Henk-Jan Westeneng, MD, Ammar Al-Chalabi, MD, PhD, Philip Van Damme, MD, PhD, Orla Hardiman, MD, PhD, Pamela J. Shaw, MD, PhD, Christopher J. McDermott, MD, PhD, Marinus J.C. Eijkemans, PhD, and Leonard H. van den Berg, MD, PhD

**Correspondence**
Dr. van Eijk
r.p.a.vaneijk-2@
umcutrecht.nl

## Abstract

Development of effective treatments for amyotrophic lateral sclerosis (ALS) has been hampered by disease heterogeneity, a limited understanding of underlying pathophysiology, and methodologic design challenges. We have evaluated 2 major themes in the design of pivotal, phase 3 clinical trials for ALS—(1) patient selection and (2) analytical strategy—and discussed potential solutions with the European Medicines Agency. Several design considerations were assessed using data from 5 placebo-controlled clinical trials (n = 988), 4 population-based cohorts (n = 5,100), and 2,436 placebo-allocated patients from the Pooled Resource Open-Access ALS Clinical Trials (PRO-ACT) database. The validity of each proposed design modification was confirmed by means of simulation and illustrated for a hypothetical setting. Compared to classical trial design, the proposed design modifications reduce the sample size by 30.5% and placebo exposure time by 35.4%. By making use of prognostic survival models, one creates a potential to include a larger proportion of the population and maximize generalizability. We propose a flexible design framework that naturally adapts the trial duration when inaccurate assumptions are made at the design stage, such as enrollment or survival rate. In case of futility, the follow-up time is shortened and patient exposure to ineffective treatments or placebo is minimized. For diseases such as ALS, optimizing the use of resources, widening eligibility criteria, and minimizing exposure to futile treatments and placebo is critical to the development of effective treatments. Our proposed design modifications could circumvent important pitfalls and may serve as a blueprint for future clinical trials in this population.

Development of treatments for amyotrophic lateral sclerosis (ALS) has been challenging and, despite considerable efforts, only riluzole has been proven to prolong survival time by 2 to 3 months.[1-3] Recently, the Treatment Research Initiative to Cure ALS has extended a collective effort among academia, patient advocacy groups, industry partners, and fundraisers to reform clinical trial design and to improve the likelihood of successful drug development.[4]

Over the years, a number of suggestions have been made to innovate the design of clinical trials in ALS.[1,3,5-7] Nevertheless, clinical trials have remained relatively conservative, especially when initiated by industry, as deviating from trial guidelines could affect regulatory acceptability. Though industry is open to fundamentally changing drug development for ALS,[4] they require amendments of the current regulatory guidelines to successfully adopt innovation in their pipelines.

In an attempt to address these concerns, we have evaluated 2 major themes in the design of pivotal, phase 3 clinical trials for ALS: (1) patient selection and (2) analytical strategy. Virtually all clinical trials impose various sets of eligibility criteria to reduce clinical heterogeneity.[5,8] This leads to the exclusion of many patients, which affects generalizability (i.e., "to whom do the results of this trial apply?").[9] Moreover, such criteria have been minimally effective in creating more homogenous trial populations.[5] In addition, funding and resources for low prevalence disorders such as ALS are limited, while costs of pivotal studies are high, requiring relatively large sample sizes and long follow-up periods to provide confirmatory evidence.[10]

To address these challenges, we illustrate the rationale for innovative design modifications in clinical trials for ALS. Our considerations were discussed with the European Medicines Agency (EMA) during a scientific advice session on June 25, 2020, and are illustrated using patient-level data from 5 placebo-controlled clinical trials (n = 988, including active and placebo patients due to lack of efficacy),[11-15] 4 population-based incidence cohorts (n = 5,100; Leuven, Sheffield, Dublin, and Utrecht), and 2,436 placebo-allocated patients from the Pooled Resource Open-Access ALS Clinical Trials (PRO-ACT) database.[5]

## Standard Protocol Approvals, Registrations, and Patient Consents

The medical ethics committee and institutional review board of the University Medical Center Utrecht approved reanalysis of the data for the purposes outlined in this work.

## Data Availability

All analyses and anonymized data will be shared by request from any qualified investigator.

# Refining Patient Eligibility: Risk-Based Patient Selection

Among the first design considerations for clinical trials are the inclusion and exclusion criteria. Depending on the study phase, selection criteria may aim to enroll a more homogeneous population, improve protocol adherence and safety, or exclude patients who are unlikely to benefit from treatment.[8,9,16] Trial eligibility is classically determined by a stepwise assessment of patient characteristics such as age or disease severity. To illustrate: trial X aims to reduce dropout and selects patients younger than 75 years with a vital capacity (VC) of ≥60%. However, fulfilling these criteria does not guarantee that the patient completes the study. For example, a patient who fails one of the criteria (e.g., patient 1, 76 years, VC 100%) may have a higher probability of completing the study as compared to a patient fulfilling all criteria (e.g., patient 2, 74 years, VC 61%). This phenomenon is explained by the fact that progression rate, prognosis, and dropout are dictated by a multivariable combination of characteristics.[5,17,18] In contrast, stepwise criteria evaluate patient characteristics univariably and ignore their combined effect on prognosis. The accuracy of patient enrollment can thus be improved by evaluating multiple characteristics simultaneously by, for example, making use of multivariable prediction models.[5]

## Risk Profile as Prognostic Summary

Prediction models are a method for summarizing multiple patient characteristics into an individual risk profile. This "prognostic summary" could serve as a targeted approach to exclude patients with an undesirable prognosis and maximize eligibility rates (i.e., the number of patients who fulfil the enrolment criteria).[5,6] Here we illustrate patient eligibility based on their (relative) survival risk profiles as estimated by the cross-validated European Network for the Cure of ALS (ENCALS) survival model.[17] Nevertheless, other prediction rules have been proposed and may be used in a similar fashion.[19] The ENCALS risk profile, hereafter referred to as "risk profile," is a weighted average of patient characteristics; its calculation is available from Dryad (eAppendix 1, doi.org/10.5061/dryad.fbg79cnv7).

In population-based cohorts, the risk profile ranges approximately from −12.0 to 0.0, with higher scores indicating a worse prognosis. Figure 1 depicts the distribution of risk profiles for 6 clinical trial cohorts (n = 3,424) compared to a population-based

incidence cohort (n = 5,100); the population ranges are available from Dryad (eAppendix 1, doi.org/10.5061/dryad.fbg79cnv7). The distribution of risk scores among clinical trial participants represents a more favorable prognosis compared to the population-based cohort. This translates to the increased survival of trial participants as compared to population-based datasets observed in previous studies.[5,8,20] Notably, patients with a risk profile of ≥−2.0 (i.e., very short survivors) are virtually non-existent in trial populations (<1.0%), while they account for 8.5% of the general ALS population. This observation is primarily driven by the delay between diagnosis and trial enrollment, a period in which most very short survivors die or become too weak to participate in a clinical trial.

### Predictive Properties of the Risk Profile for Disease Progression Rate

A key objective of inclusion criteria may be to remove non-progressing or rapidly progressing patients as (1) most treatments will require some exposure time before any effect on clinical endpoints can be detected and (2) therapeutic effects may not be efficiently measured in nonprogressing patients and would increase sample size or require a longer trial.[21] Previous studies have revealed the predictive properties of the risk profile for overall survival.[5,17] Here we illustrate that among all clinical trials reported in figure 1, the rate of decline during follow-up was strongly related to the patient's risk profile at baseline (all $p$ < 0.001; Table 1): with each unit increase in risk profile, the monthly rate of decline increased by 0.20 (95% confidence interval, 0.18–0.21) points per month. This observation may not be surprising given the predictive value of ALS Functional Rating Scale–Revised (ALSFRS-R) progression for survival time.[18] Moreover, there is a natural relationship between a poor prognosis, a short survival time, and a fast disease progression rate. Selection based on the risk profile may therefore not only affect survival and functional decline during follow-up, but might also be used to replace stepwise criteria that aim to remove fast or slow progressing patients.

### Risk-Based Patient Selection and Heterogeneity in Progression Rates

To illustrate its use as an inclusion criterion, one can define an eligibility window based on the risk profile, which can be

**Figure 1** Distribution of Risk Profiles in Clinical Trial and Population-Based Cohorts
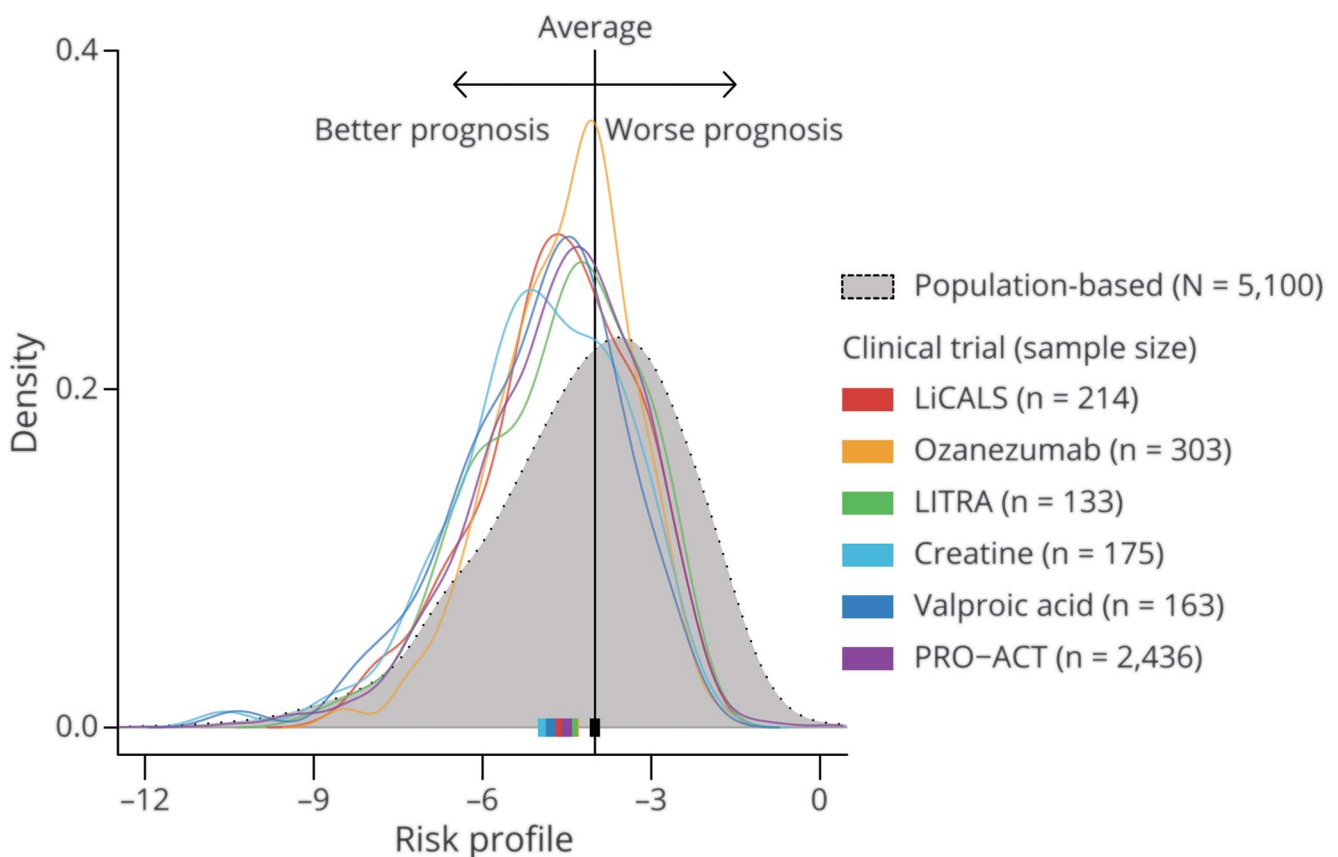


For each patient, we calculated his or her risk profile (horizontal axis), a composite score of 7 prognostic variables, which ranges from approximately −12 (very long prognosis) to 0 (very short prognosis). The distribution of the risk profiles is given as density curves, with the probability density on the vertical axis. The interpretation of the figure is similar to that of a histogram. The colors represent different trial populations; the population-based cohort is in gray. The solid squares are the population medians. A clear shift is observed in trial populations towards a better prognosis compared to a population-based cohort, which reflects the underrepresentation of patients with a poor prognosis in clinical trials. Exact ranges per cohort are available from Dryad (eAppendix 1, doi.org/10.5061/dryad.fbg79cnv7); LiCALS = lithium carbonate in patients with amyotrophic lateral sclerosis; PRO-ACT = Pooled Resource Open-Access ALS Clinical Trials.

**Table 1** Overview of the Interaction Between Disease Progression Rate and Individual Risk Profile

| Clinical trial | Year | Patients (observations), n | Rate of progression (points/mo) | Interaction risk profile–progression rate | | |
| --- | --- | --- | --- | --- | --- | --- |
| | | | | Points/mo | 95% CI | *p* Value |
| **Creatine** | 2001 | 175 (777) | −0.95 | −0.22 | −0.30 to −0.15 | <0.001 |
| **Valproic acid** | 2009 | 163 (717) | −1.01 | −0.14 | −0.21 to −0.07 | <0.001 |
| **LITRA** | 2012 | 133 (782) | −1.11 | −0.18 | −0.27 to −0.09 | <0.001 |
| **LiCALS** | 2013 | 214 (1,125) | −1.05 | −0.22 | −0.29 to −0.16 | <0.001 |
| **Ozanezumab** | 2017 | 303 (3,493) | −0.92 | −0.23 | −0.30 to −0.16 | <0.001 |
| **PRO-ACT I** | 1996–2016 | 1,688 (17,942) | −1.04 | −0.20 | −0.23 to −0.17 | <0.001 |
| **PRO-ACT II** | 1996–2016 | 748 (6,231) | −0.98 | −0.19 | −0.22 to −0.16 | <0.001 |
| **Pooled estimate:** | | | −1.01 | −0.20 | −0.21 to -0.18 | <0.001 |

Abbreviations: ALSFRS-R = ALS Functional Rating Scale–Revised; CI = confidence interval; LiCALS = lithium carbonate in patients with amyotrophic lateral sclerosis; PRO-ACT = Pooled Resource Open-Access ALS Clinical Trials.
Rate of progression is based on the ALSFRS (creatine trial and PRO-ACT II) or ALSFRS-R. The PRO-ACT database was split into 2 cohorts as some trials only measured the ALSFRS as opposed to ALSFRS-R. The interaction term indicates the increase in progression rate with each unit increase in the risk profile. Results are based on linear mixed effects models with a fixed effect for time, risk profile, and the interaction between time and risk profile. The random part contained a random slope for time and random intercept per individual. Missing information required for the calculation of the risk profile (6.63%) was imputed by multiple imputations and summarized using the Rubin rule. The pooled estimate is the meta-analyzed average across trial populations.

determined pragmatically or driven by sample size. For example, 202 patients would be required to detect a 25% reduction in ALSFRS-R slope when enrolling patients with risk profiles between −4.0 and −2.0, whereas sample size increases to 262 (+30%) when widening the window to −6.0 to −2.0.[22] The exclusion rates for each selection window are 59% and 24%, respectively. As comparison, utilizing classic criteria, for example, symptom duration ≤24 months, age ≤80 years, and VA ≥65%,[23] would result in an exclusion rate of 35% and a required sample size of 278. Although this example reveals the beneficial effect of risk-based selection on the trial's sample size, a too narrow eligibility window may restrict the trial's generalizability and compromise enrollment rates. It may therefore be advisable to define a maximum exclusion rate (say 25%) and subsequently determine which eligibility window results in the smallest sample size. To illustrate this point, in Table 2 we provide the patient characteristics of eligible patients with a risk profile −6.0 to −2.0, resulting in an exclusion rate of less than 25%. The risk-based selection results in a similar excluded population that contains the patients who are commonly omitted from trials (e.g., long disease duration, low forced VC, or rapidly progressive ALS). Of note, the final enrolled population is not solely driven by eligibility, but could also be affected by latent processes, such as patient preference, cultural differences, or physician-related factors.

## Considerations and Refinement After Regulatory Feedback

As illustrated previously,[5] the approach of using a risk profile as inclusion criterion may have potential to include a larger part of the population by better predicting prognosis, disease progression, and possibly even dropout. Nevertheless, defining the eligibility window must be carefully weighed against the reduction in sample size: can a

minimal exclusion rate be achieved without a relevant increase in sample size? The definition of a relevant loss of trial patients, and the resulting optimal eligibility window, depends on the investigator preferences and the setting (e.g., broad vs genotype-targeted treatments, or in combination with additional biomarker criteria), and may need to be fine-tuned and substantiated on a study-by-study basis. The comparability of results across trials, or evaluating shifts in the population risk profile over time, may help to further refine the eligibility window. Finally, the performance of the model must be prospectively validated, in external populations, and investigators should evaluate the model performance in both included and excluded patients. During the conduct of the trial, it would be important to carefully register intercurrent events (e.g., change in riluzole intake during follow-up) and evaluate whether and how these affect the prediction accuracy of the model.

## Optimizing Analytical Design of Survival Outcomes

In current regulatory guidelines,[3,24] an independent assessment of survival time is required to characterize efficacy in pivotal clinical trials. In the following sections, we will describe several key design considerations for ALS survival outcomes, illustrated for a hypothetical study. Because very short survivors are underrepresented in trial populations (Figure 1), the number of patients who have an event at the beginning of follow-up is considerably lower compared to the end of follow-up.[10] All calculations and simulations in the following sections are therefore based on the Weibull framework, with an increasing hazard over time and an 18-month survival of 56.1% (based on the placebo patients in PRO-ACT with a risk profile between −6.0 and −2.0).[10,25]

**Table 2** Population-Based Characteristics of Eligible Patients at Diagnosis

| Characteristics | All patients (n = 5,100) | Eligible patients (n = 3,872) | Excluded patients Slow progressors (<−6.0) (n = 793) | Fast progressors (>−2.0) (n = 435) |
|---|---|---|---|---|
| Age at diagnosis, y | 65 (11) | 65 (11) | 60 (14) | 71 (8) |
| Bulbar site of onset, n (%) | 1,618 (32) | 1,270 (33) | 94 (12) | 254 (58) |
| Diagnostic delay, months[a] | 10 (11) | 10 (7) | 29 (23) | 5 (3) |
| FVC, % predicted | 84 (22) | 85 (21) | 92 (19) | 60 (21) |
| ALSFRS-R total score | 38 (7) | 38 (6) | 40 (6) | 31 (6) |
| ΔFRS, points/mo, median (interquartile range) | −0.68 (1.03) | −0.74 (0.84) | −0.23 (0.16) | −2.86 (1.28) |
| Definite El Escorial category, n (%) | 1,330 (26) | 961 (25) | 132 (17) | 237 (54) |
| Presence of ALS-FTD, n (%) | 307 (6.0) | 216 (5.6) | 25 (3.2) | 66 (15.2) |
| 12-month survival since diagnosis, % | 66.9 | 67.3 | 87.6 | 24.5 |
| Estimated median survival, months[a] | | | | |
|   Receiving placebo | 18.7 | 18.6 | 37.4 | 6.1 |
|   Receiving treatment (HR 0.63) | 27.4 (+8.7) | 27.2 (+8.6) | 52.0 (+14.6) | 9.8 (+3.7) |

Abbreviations: ΔFRS = ALSFRS-R score – 48/diagnostic delay; ALSFRS-R = ALS Functional Rating Scale–Revised; FTD = frontotemporal dementia; FVC = forced vital capacity; HR = hazard ratio.
Patients <−6.0 are patients with an excellent prognosis; patients >−2.0 are patients with a poor prognosis. Data are mean (SD) or % (n) unless noted otherwise.
[a] Median survival was estimated from a parametric Weibull survival model. "Receiving placebo" is the median survival time of the cohort with no benefit of treatment, whereas "Receiving treatment" reflects a hypothetical scenario that illustrates how median survival improves when treatment is effective.

## Optimizing Time Requirements: Extended Randomized Follow-up

Optimizing the use of time can considerably improve the power of survival outcomes to detect treatment responses.[26] Because the number of events (e.g., deaths) drives their power,[27,28] solely extending follow-up time, while keeping the sample size constant, can increase the power to detect treatment effects. An important consideration is therefore to extend the randomized follow-up for early-enrolled patients and, as a consequence, generate more events within the same timeframe.[10] To illustrate: an 18-month trial (n = 300) with a 12-month enrollment period could extend the follow-up for early-enrolled patients, while retaining an 18-month follow-up for late-enrolled patients, increasing power from 68.5% to 84.6% to detect a hazard ratio (HR) of 0.63.[28,29] The extended follow-up can be implemented with minimal effort and could result in important reductions in sample size and costs.[10,26] From a patient perspective, the drawback of such an approach is the uncertainty about the length of trial participation and the potentially longer time on placebo if enrolled early. These limitations could be mitigated by implementing interim analyses (illustrated below), or by using a hybrid approach, with a maximum follow-up period for early enrolled patients. An example of the latter can be found in the dexpramipexole study.[23]

## Accuracy of a Priori Assumptions and Trial Design

Designing clinical trials with time-to-event endpoints requires a considerable number of assumptions; for example, we made

assumptions above about the survival pattern (Weibull), survival probability (56.1%), and enrollment period (12 months). Despite the fact that these assumptions can be estimated using historical data, an inherent risk of inaccuracy remains. Making inaccurate a priori assumptions at the design stage (i.e., misspecification) can have detrimental consequences for a trial's ability to draw definite conclusions.[30,31] If the observed 18-month survival is 50.0% or 60.0% instead of 56.1%, the required sample size fluctuates between 237 and 286 patients (a difference of up to 21%; data available from Dryad, eAppendix 1, doi.org/10.5061/dryad.fbg79cnv7). As a result, inaccuracy in design assumptions could significantly over- or underpower clinical trials and unnecessarily expose patients to harmful or ineffective treatments.

## Event-Driven Trial Design as Guard Against Misspecification

The required sample size for time-to-event outcomes is calculated as the required number of events divided by the probability of an event.[29] The probability of an event depends on various assumptions (e.g., survival distribution, dropout rate, and enrollment period), whereas the number of events depends solely on the hypothesized treatment effect (HR), allocation ratio, power, and type 1 error.[27,28] In other words, irrespective of the follow-up duration, survival distribution, or enrollment rate, as long as the required number of events is obtained, the trial will reach its designated power. This is a useful feature of time-to-event endpoints and it has been

suggested, therefore, to run the trial until a prespecified number of events is reached (i.e., event-driven or information-based trials) rather than concluding the trial after a fixed duration of follow-up (i.e., duration-based trials).[32-34] The major benefit of the event-driven approach is that the accuracy of the event probability becomes irrelevant. The major drawback is that the exact trial duration is unknown (i.e., when is the target number of events reached?) and cannot be fixed from the outset.[32] Nevertheless, using various projections about the underlying assumptions, the uncertainty in trial duration can be estimated and used as a guide by investigators, or to inform patients about the expected individual follow-up time.

## Group-Sequential Interim Analyses

The last design consideration is to implement an interim analysis scheme to stop the trial early because of either inefficacy (i.e., futility or harm) or efficacy (i.e., superiority). Implementing interim analyses could considerably increase a trial's efficiency and reduce the exposure of patients to ineffective treatments or placebo.[7,35] Similar to the final analysis, the timing of interim analyses can be guided by the number of events. Implementing an interim analysis scheme does, however, increase the required number of events to account for multiplicity. For example, incorporating a single interim analysis after the occurrence of 60% of the events requires 2.9% more events compared to a design without interim analyses (using a conservative O'Brien-Fleming type alpha-spending function).[36] The expected number of events required to reach a conclusion is, however, 25.6% less if the treatment is futile, and 13.9% less if the treatment is superior. Thus, if a trial cannot be stopped early, slightly more events are needed, but on average fewer events are required to reach a conclusion compared to a design without interim analyses.[7]

## Illustration of Event-Driven, Group-Sequential Trial

Finally, by means of simulation, we will illustrate the above-mentioned design considerations for a hypothetical study. The study targets a total of 153 events to detect an HR of 0.63 with 80% power and a 1-sided α of 2.5%. In this setting, an HR of 0.63 translates to a median survival increase after enrollment from 19.7 to 24.8 months (+5.1 months); in Table 2, the median survival estimates are provided for other populations. A single interim analysis will occur after 60% of the required events. Assuming an enrollment period of 12 months, it can be estimated that 278 patients are required (139 active vs 139 placebo) to reach the target number of events within 30 months after first enrolled patient.[10] The trial setting was simulated under both the null (i.e., HR 1) and the alternative (i.e., HR 0.63) hypothesis, where each simulated trial was run until an interim decision was reached or until 153 events occurred. Based on 100,000 simulations, the 1-sided type I error was 2.45%, whereas the empirical power was 80.4%, confirming the validity of the design.

**Figure 2** Expected Trial Duration Under the Null and Alternative Hypothesis
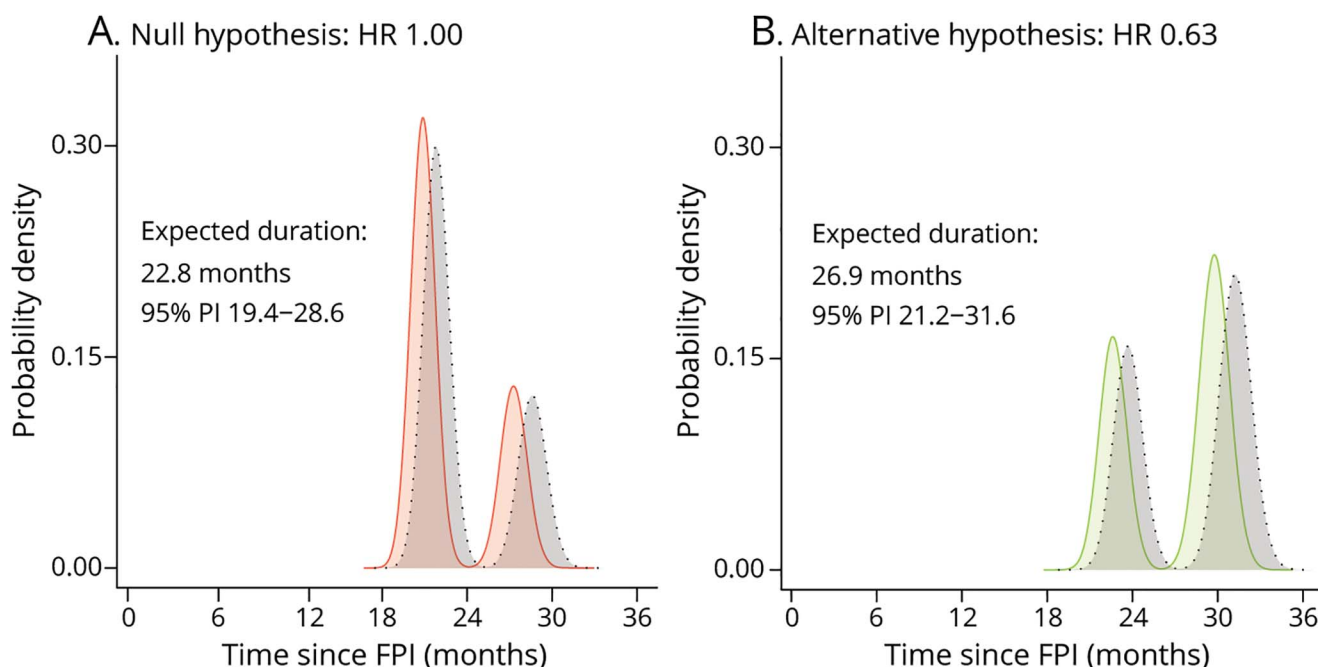


Distribution of simulated trial duration under the null hypothesis (H₀, i.e., hazard ratio [HR] 1, panel A) and alternative hypothesis (H₁, i.e., HR 0.63, panel B). The vertical axis reflects the probability density; its interpretation is similar to a histogram. Due to the lack of a treatment benefit under H₀, the required number of events occurs more rapidly and the average duration is shorter compared to H₁. The peaks reflect the average analysis time points; the last peak in panel B is centered around 30 months (i.e., when all assumptions hold). In gray is a scenario when the survival probability in the placebo arm is better than expected (60.0% instead of 56.1%). FPI = first patient first visit; PI = percentile interval of the empirical trial duration ranging from the 2.5th to the 97.5th percentile.

**Table 3** Classical vs Event-Driven, Group-Sequential Design of a Hypothetical Clinical Trial for Amyotrophic Lateral Sclerosis

| Parameter | Classical design | Event-driven, GSD design |
|---|---|---|
| **Planned sample size** | 400 | 278 (−30.5%) |
| **Required number of events** | 148 | 153 (+2.9%) |
| **Empirical power, %** | 79.5 | 80.4 |
| **Trial duration, mo** | | |
| HR 1.00 ($H_0$) | 35.3 | 22.8 (−35.4%) |
| HR 0.63 ($H_1$) | 35.3 | 26.9 (−23.8%) |
| **Total placebo exposure, person-months** | | |
| HR 1.00 ($H_0$) | 3,012 | 1,947 (−35.4%) |
| HR 0.63 ($H_1$) | 3,012 | 2,249 (−25.3%) |
| **Average placebo exposure, mo** | | |
| HR 1.00 ($H_0$) | 15.1 | 14.0 (−7.0%) |
| HR 0.63 ($H_1$) | 15.1 | 16.2 (+7.4%) |
| **Follow-up costs (million $)** | | |
| HR 1.00 ($H_0$) | 2.94 | 1.90 (−35.2%) |
| HR 0.63 ($H_1$) | 3.02 | 2.18 (−27.7%) |

Abbreviations: GSD = group-sequential design; HR = hazard ratio.
Classical design illustrates a 1:1 randomized trial with a fixed follow-up of 18 months and 80% power to detect an HR of 0.63 with an expected survival probability in the placebo arm of 56.1% after 18 months. The event-driven GSD design extends randomized follow-up for early enrolled patients, stops when the target number of events is reached, and incorporates one interim analysis after 60% of the required number of events during which the trial can be stopped for either efficacy or futility. We assumed a constant enrollment rate of 278 patients over 12 months. Trial duration is defined as the time from first patient first visit to last patient last visit (Figure 2). The total placebo exposure is defined as the sum of all follow-up times for placebo patients. The average placebo exposure is the mean follow-up time for placebo patients. Follow-up costs are defined as the sum of individual follow-up costs for all patients (data available from Dryad, eAppendix 1, doi. org/10.5061/dryad.fbg79cnv7).

The stopping times of each simulated trial are provided in Figure 2. The peaks reflect the average (interim) analysis time points. The last peak in Figure 2B is centered on the 30 months mark; this reflects the expected trial duration if all assumptions hold, including the treatment effect, and the trial cannot be stopped early. If the treatment is futile compared to placebo (e.g., HR 1), the target number of events is reached sooner and the trial can be stopped earlier. Table 3 provides a comparison with a classical design; the illustrated design results in considerable reductions in sample size, trial duration, total placebo exposure time, and costs. Individual patients, on average, may be exposed longer (+7.4%) to placebo compared to a fixed design when treatment is effective, but shorter when treatment is ineffective (−7.0%). The event-driven design naturally adapts the trial duration to reach the target number of events when one of the design assumptions is inaccurate (e.g., underestimation of the survival probability, illustrated in

gray, Figure 2). This underscores the resilience of event-driven designs against misspecification of assumptions and the assurance that power will be attained.

## Considerations and Refinement After Regulatory Feedback

An effect on only function without an effect on survival has so far not been accepted by the EMA. This may change in the future if the association between survival and function becomes more firmly established (e.g., when a study reveals how a treatment benefit on ALSFRS-R is reflected in life expectancy). For current pivotal studies, it remains therefore critical to obtain adequate information on survival time. Group-sequential methodology results, on average, in smaller and shorter trials, although the number of interim analyses should be justified and not large. A helpful guide could be to determine whether sufficient information on safety, important subgroups, and observational time per patient would be available if a trial is stopped early. To illustrate: if an interim analysis in Figure 2 were performed after 10% of the events, the average follow-up time per patient would only be 4.9 months (vs 14.4 months at 60% of the events). Should a study be stopped early for efficacy, existing safety datasets and means of collecting additional safety data (e.g., utilizing postapproval studies) would have to be discussed with regulatory authorities.

## Discussion

Modifications in the design of clinical trials may not only result in large efficiency gains and reductions in costs, but also widen eligibility criteria and minimize patients' exposure time to ineffective treatments or placebo. Considering the high futility rates in previous ALS clinical trials, and the considerable number of promising treatments, it is critical that the design of future studies is optimized. Our proposed design modifications could provide an important step forwards, where the concepts may serve as a blueprint for future clinical trials with time-to-event endpoints in ALS.

ALS clinical trials are primarily affected by clinical and pathophysiologic heterogeneity, which complicates the detection of treatment effects. Here we quantified the prognostic heterogeneity in clinical trials by using individual risk profiles, and subsequently reduced the observed heterogeneity by using the risk profile as the eligibility criterion.[5,37] Nevertheless, the risk profile may also be used to study between-trial differences and improve between-study comparability, whereas its distribution can provide insight into the generalizability of trial results and help to define the label for market authorization.[38] For trial design specifically, risk profiles can be used to improve randomization,[6] explore risk-based subgroup analyses,[38] or increase statistical power as a covariate in the final analysis.[39] Although the risk profile in our study was based on the ENCALS model, the methodology is not limited to a specific model. Continuously optimizing the available

prediction rules remains of importance to improve future study design.

As ALS trials are becoming increasingly complex with the arrival of platform trials (e.g., NCT04297683 and NCT04302870) and genotype-targeted studies (e.g., NCT03626012), uncertainty at the design stage may play a progressively more important role. Event-driven or information-based designs may serve as a relatively straightforward strategy to ameliorate the consequences of inaccurate design assumptions, and protect the trial's (or substudies') ability to draw definite conclusions. The main limitation of event-driven designs is the relatively unknown trial duration, both for investigators and patients, and the potentially longer exposure to placebo when enrolled early. This is balanced by the guarantee in statistical power, where simulations can provide extensive insights into the variability between best- and worst-case scenarios and help to refine sample size calculations, also by implementation of interim analyses.

The proposed design can be further refined by integrating clinical outcomes with survival time.[18,40] These composite endpoints could provide enhanced insights into the expected treatment effect, reduce the effects of (informative) missing data, and may further optimize decision-making about whether to stop a trial early or continue it. Moreover, the position of the regulator may change in the future with a shift towards intermediate endpoints such as ALSFRS-R or respiratory function. Group-sequential designs, extended randomized follow-up, and information-based design may be of similar value in these settings, and a continued dialogue with regulatory bodies will be essential to bring innovation to clinical trials in ALS.

Ultimately, trial design is a dynamic process, where obtaining input from patients, industry, regulators, and funders will be fundamental to the outline of a definite roadmap and to achieving wide-scale adoption. Introducing alternative concepts for the design of clinical trials in ALS may circumvent important pitfalls encountered by classical designs and speed up the development of effective therapy for this debilitating disease.

## Study Funding

## Disclosure

## Publication History

### Appendix Authors

| Name | Location | Contribution |
|---|---|---|
| Ruben P.A. van Eijk, MD, PhD | UMC Utrecht, the Netherlands | Designed and conceptualized study; analyzed the data; drafted the manuscript for intellectual content |
| Stavros Nikolakopoulos, PhD | UMC Utrecht, the Netherlands | Designed and conceptualized study; interpreted the data; revised the manuscript for intellectual content |
| Kit C.B. Roes, PhD | Radboud UMC, Nijmegen, the Netherlands | Designed and conceptualized study; interpreted the data; revised the manuscript for intellectual content |
| Lindsay Kendall, MSc | GlaxoSmithKline R&D, Stevenage, UK | Major role in the acquisition of data; revised the manuscript for intellectual content |
| Steve S. Han, MD, PhD | GlaxoSmithKline R&D, Upper Providence, PA | Major role in the acquisition of data; revised the manuscript for intellectual content |
| Arseniy Lavrov, MD, PhD | GlaxoSmithKline R&D, Middlesex, UK | Major role in the acquisition of data; revised the manuscript for intellectual content |
| Noam Epstein, MD | GlaxoSmithKline R&D, Upper Providence, PA | Major role in the acquisition of data; revised the manuscript for intellectual content |
| Tessa Kliest, MSc | UMC Utrecht, the Netherlands | Major role in the acquisition of data; revised the manuscript for intellectual content |
| Adriaan D. de Jongh, MD | UMC Utrecht, the Netherlands | Interpreted the data; revised the manuscript for intellectual content |
| Henk-Jan Westeneng, MD | UMC Utrecht, the Netherlands | Interpreted the data; revised the manuscript for intellectual content |
| Ammar Al-Chalabi, MD, PhD | King's College Hospital, London, UK | Major role in the acquisition of data; revised the manuscript for intellectual content |
| Philip Van Damme, MD, PhD | University Hospitals Leuven, Belgium | Major role in the acquisition of data; revised the manuscript for intellectual content |
| Orla Hardiman, MD, PhD | Beaumont Hospital, Dublin, Ireland | Major role in the acquisition of data; revised the manuscript for intellectual content |
| Pamela J. Shaw, MD, PhD | Sheffield Institute for Translational Neuroscience, UK | Major role in the acquisition of data; revised the manuscript for intellectual content |
| Christopher J. McDermott, MD, PhD | Sheffield Institute for Translational Neuroscience, UK | Major role in the acquisition of data; revised the manuscript for intellectual content |

Continued

## Appendix *(continued)*

| Name | Location | Contribution |
|------|----------|-------------|
| **Marinus J.C. Eijkemans, PhD** | UMC Utrecht, the Netherlands | Designed and conceptualized study; interpreted the data; revised the manuscript for intellectual content |
| **Leonard H. van den Berg, MD, PhD** | UMC Utrecht, the Netherlands | Designed and conceptualized study; interpreted the data; revised the manuscript for intellectual content |

## References

1. Mitsumoto H, Brooks BR, Silani V. Clinical trials in amyotrophic lateral sclerosis: why so many negative trials and how can trials be improved? *Lancet Neurol.* 2014;13(11): 1127-1138.
2. Chio A, Mazzini L, Mora G. Disease-modifying therapies in amyotrophic lateral sclerosis. *Neuropharmacology.* 2020;167:107986.
3. van den Berg LH, Sorenson E, Gronseth G, et al. Revised Airlie House consensus guidelines for design and implementation of ALS clinical trials. *Neurology.* 2019; 92(14):e1610-e1623.
4. van Eijk RPA, Kliest T, McDermott CJ, et al. TRICALS: creating a highway toward a cure. *Amyotroph Lateral Scler Frontotemporal Degener.* 2020(7-8);21:496-501.
5. van Eijk RPA, Westeneng HJ, Nikolakopoulos S, et al. Refining eligibility criteria for amyotrophic lateral sclerosis clinical trials. *Neurology.* 2019;92(5):e451-e460.
6. Berry JD, Taylor AA, Beaulieu D, et al. Improved stratification of ALS clinical trials using predicted survival. *Ann Clin Transl Neurol.* 2018;5(4):474-485.
7. van Eijk RPA, Nikolakopoulos S, Ferguson TA, Liu D, Eijkemans MJC, van den Berg LH. Increasing the efficiency of clinical trials in neurodegenerative disorders using group sequential trial designs. *J Clin Epidemiol.* 2018;98:80-88.
8. Chio A, Canosa A, Gallo S, et al. ALS clinical trials: do enrolled patients accurately represent the ALS population? *Neurology.* 2011;77(15):1432-1437.
9. Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?" *Lancet.* 2005;365(9453):82-93.
10. 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(12):1331-1337.
11. Groeneveld GJ, Veldink JH, van der Tweel I, et al. A randomized sequential trial of creatine in amyotrophic lateral sclerosis. *Ann Neurol.* 2003;53(4):437-445.
12. UKMND-LiCALS Study Group, Morrison KE, Dhariwal S, Hornabrook R, et al. Lithium in patients with amyotrophic lateral sclerosis (LiCALS): a phase 3 multicentre, randomised, double-blind, placebo-controlled trial. *Lancet Neurol.* 2013;12(4): 339-345.
13. Piepers S, Veldink JH, de Jong SW, et al. Randomized sequential trial of valproic acid in amyotrophic lateral sclerosis. *Ann Neurol.* 2009;66(2):227-234.
14. Verstraete E, Veldink JH, Huisman MH, et al. Lithium lacks effect on survival in amyotrophic lateral sclerosis: a phase IIb randomised sequential trial. *J Neurol Neurosurg Psychiatry.* 2012;83(5):557-564.
15. Meininger V, Genge A, van den Berg LH, et al. Safety and efficacy of ozanezumab in patients with amyotrophic lateral sclerosis: a randomised, double-blind, placebo-controlled, phase 2 trial. *Lancet Neurol.* 2017;16(3):208-216.
16. Atassi N, Yerramilli-Rao P, Szymonifka J, et al. Analysis of start-up, retention, and adherence in ALS clinical trials. *Neurology.* 2013;81(15):1350-1355.
17. Westeneng HJ, Debray TPA, Visser AE, et al. Prognosis for patients with amyotrophic lateral sclerosis: development and validation of a personalised prediction model. *Lancet Neurol.* 2018;17(5):423-433.
18. van Eijk RPA, Eijkemans MJC, 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.
19. Xu L, He B, Zhang Y, et al. Prognostic models for amyotrophic lateral sclerosis: a systematic review. *J Neurol.* Epub 2021 Mar 10.
20. Hardiman O, Al-Chalabi A, Brayne C, et al. The changing picture of amyotrophic lateral sclerosis: lessons from European registers. *J Neurol Neurosurg Psychiatry.* 2017; 88(7):557-563.
21. de Carvalho M, Swash M. Can selection of rapidly progressing patients shorten clinical trials in amyotrophic lateral sclerosis? *Arch Neurol.* 2006;63(4):557-560.
22. Ard MC, Edland SD. Power calculations for clinical trials in Alzheimer's disease. *J Alzheimers Dis.* 2011;26(suppl 3):369-377.
23. Cudkowicz ME, van den Berg LH, Shefner JM, et al. Dexpramipexole versus placebo for patients with amyotrophic lateral sclerosis (EMPOWER): a randomised, double-blind, phase 3 trial. *Lancet Neurol.* 2013;12(11):1059-1067.
24. Andrews JA, Bruijn LI, Shefner JM. ALS drug development guidances and trial guidelines: consensus and opportunities for alignment. *Neurology.* 2019;93(2):66-71.
25. Wu J. *Statistical Methods for Survival Trial Design: With Applications to Cancer Clinical Trials Using R.* Taylor & Francis; 2018.
26. Bateman RJ, Benzinger TL, Berry S, et al. The DIAN-TU Next Generation Alzheimer's prevention trial: adaptive design and disease progression model. *Alzheimers Dement.* 2017;13(1):8-19.
27. Freedman LS. Tables of the number of patients required in clinical trials using the logrank test. *Stat Med.* 1982;1(2):121-129.
28. Schoenfeld DA, Richter JR. Nomograms for calculating the number of patients needed for a clinical trial with survival as an endpoint. *Biometrics.* 1982;38(1):163-170.
29. Abel UR, Jensen K, Karapanagiotou-Schenkel I, Kieser M. Some issues of sample size calculation for time-to-event endpoints using the Freedman and Schoenfeld formulas. *J Biopharm Stat.* 2015;25(6):1285-1311.
30. Nikolakopoulos S, Roes KC, van der Lee JH, van der Tweel I. Sample size calculations in pediatric clinical trials conducted in an ICU: a systematic review. *Trials.* 2014;15:274.
31. Tavernier E, Giraudeau B. Sample size calculation: inaccurate a priori assumptions for nuisance parameters can greatly affect the power of a randomized controlled trial. *PLoS One.* 2015;10(7):e0132578.
32. Lachin JM. Maximum information designs. *Clin Trials* 2005;2(5):453-464.
33. Freidlin B, Othus M, Korn EL. Information time scales for interim analyses of randomized clinical trials. *Clin Trials.* 2016;13(4):391-399.
34. Kim K, Boucher H, Tsiatis AA. Design and analysis of group sequential logrank tests in maximum duration versus information trials. *Biometrics.* 1995;51(3): 988-1000.
35. Groeneveld GJ, van der Tweel I, Wokke JH, van den Berg LH. Sequential designs for clinical trials in amyotrophic lateral sclerosis. *Amyotroph Lateral Scler Other Mot Neuron Disord.* 2004;5(4):202-207.
36. Kim k, DeMets DL. Design and analysis of group sequential tests based on the type I error spending rate function. *Biometrika.* 1987;74(1):149-154.
37. Hardiman O, van den Berg LH. Edaravone: a new treatment for ALS on the horizon? *Lancet Neurol.* 2017;16(7):490-491.
38. Ioannidis JP, Lau J. Heterogeneity of the baseline risk within patient populations of clinical trials: a proposed evaluation algorithm. *Am J Epidemiol.* 1998;148(11): 1117-1126.
39. Kahan BC, Jairath V, Dore CJ, Morris TP. The risks and rewards of covariate adjustment in randomized trials: an assessment of 12 outcomes from 8 studies. *Trials.* 2014;15:139.
40. 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(3):162-168.

# The AAN is at Your Side

When you're in the office, the AAN is at your side. The AAN is your #1 resource to support you and your care team. Whether it's resources to help you and your staff provide the best care for your patients, ensure proper reimbursement, or maximize practice performance, the AAN is at your side. Access these resources today at *AAN.com/practice.*