Challenging the Established Order: Innovating Clinical Trials for Amyotrophic Lateral Sclerosis

Ruben P.A. van Eijk,1,2 Stavros Nikolakopoulos,2 Kit C.B. Roes,3 Lindsay Kendall,4 Steve S. Han,5,6 Arseniy Lavrov,7,8 Noam Epstein,6 Tessa Kliest,1 Adriaan D. de Jongh,1 Henk-Jan Westeneng,1 Ammar Al-Chalabi,9,10 Philip Van Damme,11,12 Orla Hardiman,13,14 Pamela J. Shaw,15 Christopher J. McDermott,15 Marinus J.C. Eijkemans,2 Leonard H. van den Berg1

This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (CC BY), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Neurology® Published Ahead of Print articles have been peer reviewed and accepted for publication. This manuscript will be published in its final form after copyediting, page composition, and review of proofs. Errors that could affect the content may be corrected during these processes.
(1) Department of Neurology, UMC Utrecht Brain Center, University Medical Center Utrecht, Utrecht, the Netherlands.

(2) Biostatistics & Research Support, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, the Netherlands.

(3) Department of Health Evidence, Section Biostatistics, Radboud Medical Centre Nijmegen, the Netherlands.

(4) Biostatistics, GlaxoSmithKline R&D, Stevenage, UK.

(5) Neurosciences, Takeda Pharmaceuticals, Cambridge, USA.

(6) Discovery Medicine, GlaxoSmithKline R&D, Upper Providence, USA.


(8) Clinical Translational Medicine, Future Pipeline Discovery, GlaxoSmithKline R&D, Middlesex, UK.

(9) King's College London, London, Maurice Wohl Clinical Neuroscience Institute and United Kingdom Dementia Research Institute Centre, Department of Basic and Clinical Neuroscience, UK.

(10) Department of Neurology, King's College Hospital, London, UK.

(11) Department of Neurosciences, Laboratory for Neurobiology, KU Leuven and Center for Brain & Disease Research, VIB, Leuven Brain Institute, Leuven, Belgium.

(12) Department of Neurology, University Hospitals Leuven, Leuven, Belgium.

(13) Department of Neurology, National Neuroscience Centre, Beaumont Hospital, Dublin, Ireland.

(14) FutureNeuro SFI Research Centre, Royal College of Surgeons in Ireland, Dublin, Ireland

(15) Department of Neuroscience, University of Sheffield, Sheffield Institute for Translational Neuroscience, Sheffield, UK.
Corresponding author
Ruben P.A. van Eijk
Email: r.p.a.vaneijk-2@umcutrecht.nl

Title character count: 89/140
Word count: 3,497/3,500
Abstract: 248/350
Tables: 3
Figures: 2
References: 40/40

Study Funding
This study was funded by the Dutch ALS Foundation, UK Motor Neurone Disease Association (MNDA) and ALS Liga Belgium. It is in part an EU Joint Programme - Neurodegenerative Disease Research (JPND) project - www.jpnd.eu (United Kingdom, Medical Research Council (MR/L501529/1; MR/R024804/1)). This study represents independent research, in part funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King’s College London. PJS is supported as an NIHR Senior Investigator. CJM and PJS are supported by the NIHR Sheffield Biomedical Research Centre (BRC).

Disclosure
L. Kendall and N. Epstein are current, and S.S. Han and A. Lavrov are former employees of GlaxoSmithKline and held shares in the company. The other authors declare that they have no conflict of interest.
ABSTRACT

Development of effective treatment for amyotrophic lateral sclerosis (ALS) has been hampered by disease heterogeneity, a limited understanding of underlying pathophysiology and methodological design challenges. Here we have evaluated two 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 (EMA).

Several design considerations were assessed using data from five placebo-controlled clinical trials (N = 988), four population-based cohorts (N = 5,100), and 2,436 placebo-allocated patients from the 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 which naturally adapts the trial duration when inaccurate assumptions are made at the design stage such as the 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 the 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 a blueprint for future clinical trials in this population.
INTRODUCTION

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 two to three months.\textsuperscript{1-3} Recently, the Treatment Research Initiative to Cure ALS (TRICALS) 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.\textsuperscript{4}

Over the years, a number of suggestions have been made to innovate the design of clinical trials in ALS.\textsuperscript{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,\textsuperscript{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 two 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.\textsuperscript{5, 8} This leads to the exclusion of many patients, which affects generalizability (i.e. “to whom do the results of this trial apply?”).\textsuperscript{9} Moreover, such criteria have been minimally effective in creating more homogenous trial populations.\textsuperscript{5} In addition, funding and resources for low prevalent 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.\textsuperscript{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 procedure on June 25, 2020, and are illustrated using patient-level data from five placebo-controlled clinical trials (N = 988, including active and placebo patients due to lack of efficacy), \(^{11-15}\) four population-based incidence cohorts (N = 5,100; Leuven, Sheffield, Dublin and Utrecht), and 2,436 placebo-allocated patients from the 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 re-analysis 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.

**1. 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 step-wise 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 of age and with a vital capacity (VC) of \(\geq 60\%\). 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, step-wise criteria evaluate patient characteristics *univariably* and ignore their combined impact 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.\textsuperscript{5}

1.1. The 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).\textsuperscript{5,6} Here we illustrate patient eligibility based on their (relative) survival risk profiles as estimated by the cross-validated ENcals survival model.\textsuperscript{17} Nevertheless, other prediction rules have been proposed and may be used in a similar fashion.\textsuperscript{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; https://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 six clinical trial cohorts (n = 3,424) compared to a population-base incidence cohort (n = 5,100); the population ranges are available from Dryad (eAppendix 1; https://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.\textsuperscript{5,8,20} Notably, patients with a risk profile of $\geq -2.0$ (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.
1.2. 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 non-progressing patients, and would increase sample size or require a longer trial.\textsuperscript{21} Previous studies have revealed the predictive properties of the risk profile for overall survival.\textsuperscript{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% CI 0.18 – 0.21) points per month. This observation may not be surprising given the predictive value of ALSFRS-R progression for survival time.\textsuperscript{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 step-wise criteria that aim to remove fast or slow progressing patients.

1.3. Risk-based patient selection and heterogeneity in progression rates

To illustrate its use as inclusion criterion, one can define an eligibility window based on the risk profile, which can be 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].\textsuperscript{22} The exclusion rates for each selection window are 59\% and 24\%, respectively. As comparison, utilizing classical criteria, e.g. symptom duration $\leq$24 months, age $\leq$80 years and vital capacity $\geq$65\%,\textsuperscript{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 FVC and/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.

1.4. Considerations and refinement after regulatory feedback
As illustrated previously, 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.
2. Optimizing analytical design of survival outcomes

In current regulatory guidelines,\textsuperscript{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. Since very short survivors are under-represented in trial populations (Figure 1), the number of patients that have an event at the beginning of follow-up is considerably lower compared to the end of follow-up.\textsuperscript{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).\textsuperscript{10, 25}

2.1. Optimizing time-requirements: extended randomized follow-up

Optimizing the use of time can considerably improve the power of survival outcomes to detect treatment responses.\textsuperscript{26} Since the number of events (e.g. deaths) drives their power,\textsuperscript{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.\textsuperscript{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.\textsuperscript{28, 29} The extended follow-up can be implemented with minimal effort and could result in important reductions in sample size and costs.\textsuperscript{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.\textsuperscript{23}

\textbf{2.2. Accuracy of a priori assumptions and trial design}

Designing clinical trials with time-to-event endpoints requires a considerable number of assumptions; e.g. above we made assumptions 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.\textsuperscript{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; https://doi.org/10.5061/dryad.fbg79cnv7). As a result, inaccuracy in design assumptions could significantly over- or under-power clinical trials and unnecessarily expose patients to harmful or ineffective treatments.

\textbf{2.3. Event-driven trial design as guard against misspecification}

The required sample size for time-to-event outcomes is calculated as the \textit{required number of events} divided by the \textit{probability of an event}.\textsuperscript{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.\textsuperscript{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 pre-specified 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).\textsuperscript{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. 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.

2.4. Group-sequential interim analyses

The last design consideration is to implement an interim analysis scheme to stop the trial early because of either ineffectiveness (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. 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). 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.

2.5. 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 a HR of 0.63 with 80% power and a one-sided alpha of 2.5%. In this setting, a 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. 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 one-sided type I error was 2.45%, whereas the empirical power was 80.4%, confirming the validity of the design.

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. Importantly, if the treatment is futile compared to placebo (e.g. HR of 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%). Importantly, 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 grey, Figure 2). This underscores the resilience of event-driven designs against misspecification of assumptions and the assurance that power will be attained.

2.6. 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 post-approval 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 the patient’s 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 pathophysiological 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. 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.\textsuperscript{38} For trial design specifically, risk profiles can be used to improve randomization,\textsuperscript{6} explore risk-based subgroup analyses,\textsuperscript{38} or increase statistical power as a covariate in the final analysis.\textsuperscript{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, therefore, 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), the uncertainty at the design stage may play progressively a 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 sub-studies’) 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.\textsuperscript{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.
## APPENDIX 1: AUTHORS

<table>
<thead>
<tr>
<th>Name</th>
<th>Location</th>
<th>Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ruben P.A. van Eijk, MD, PhD</td>
<td>UMC Utrecht, Utrecht,</td>
<td>Designed and conceptualized study; analyzed the data; drafted the manuscript for intellectual content</td>
</tr>
<tr>
<td></td>
<td>The Netherlands</td>
<td></td>
</tr>
<tr>
<td>Stavros Nikolakopoulos, PhD</td>
<td>UMC Utrecht, Utrecht,</td>
<td>Designed and conceptualized study; Interpreted the data; revised the manuscript for intellectual content</td>
</tr>
<tr>
<td></td>
<td>The Netherlands</td>
<td></td>
</tr>
<tr>
<td>Kit C.B. Roes, PhD</td>
<td>Radboud UMC, Nijmegen, The</td>
<td>Designed and conceptualized study; Interpreted the data; revised the manuscript for intellectual content</td>
</tr>
<tr>
<td></td>
<td>Netherlands</td>
<td></td>
</tr>
<tr>
<td>Lindsay Kendall, MSc</td>
<td>GlaxoSmithKline R&amp;D, Stevenage, UK</td>
<td>Major role in the acquisition of data; revised the manuscript for intellectual content</td>
</tr>
<tr>
<td>Steve S. Han, MD, PhD</td>
<td>GlaxoSmithKline R&amp;D, Upper Providence, USA</td>
<td>Major role in the acquisition of data; revised the manuscript for intellectual content</td>
</tr>
<tr>
<td>Arseniy Lavrov, MD, PhD</td>
<td>GlaxoSmithKline R&amp;D, Middlesex, UK</td>
<td>Major role in the acquisition of data; revised the manuscript for intellectual content</td>
</tr>
<tr>
<td>Noam Epstein, MD</td>
<td>GlaxoSmithKline R&amp;D, Upper Providence, USA</td>
<td>Major role in the acquisition of data; revised the manuscript for intellectual content</td>
</tr>
<tr>
<td>Name</td>
<td>Institution</td>
<td>Contributions</td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>--------------------------------------------</td>
<td>---------------------------------------------------------</td>
</tr>
<tr>
<td>Tessa Kliest, MSc</td>
<td>UMC Utrecht, Utrecht, The Netherlands</td>
<td>Major role in the acquisition of data; revised the manuscript for intellectual content</td>
</tr>
<tr>
<td>Adriaan D. de Jongh, MD</td>
<td>UMC Utrecht, Utrecht, The Netherlands</td>
<td>Interpreted the data; revised the manuscript for intellectual content</td>
</tr>
<tr>
<td>Henk-Jan Westeneng, MD</td>
<td>UMC Utrecht, Utrecht, The Netherlands</td>
<td>Interpreted the data; revised the manuscript for intellectual content</td>
</tr>
<tr>
<td>Ammar Al-Chalabi, MD, PhD</td>
<td>King’s College Hospital, London, UK</td>
<td>Major role in the acquisition of data; revised the manuscript for intellectual content</td>
</tr>
<tr>
<td>Philip Van Damme, MD, PhD</td>
<td>University Hospitals Leuven, Leuven, Belgium</td>
<td>Major role in the acquisition of data; revised the manuscript for intellectual content</td>
</tr>
<tr>
<td>Orla Hardiman, MD, PhD</td>
<td>Beaumont Hospital, Dublin, Ireland</td>
<td>Major role in the acquisition of data; revised the manuscript for intellectual content</td>
</tr>
<tr>
<td>Pamela J. Shaw, MD, PhD</td>
<td>Sheffield Institute for Translational Neuroscience, Sheffield, UK</td>
<td>Major role in the acquisition of data; revised the manuscript for intellectual content</td>
</tr>
<tr>
<td>Christopher J. McDermott, MD, PhD</td>
<td>Sheffield Institute for Translational Neuroscience, Sheffield, UK</td>
<td>Major role in the acquisition of data; revised the manuscript for intellectual content</td>
</tr>
<tr>
<td>Author(s)</td>
<td>Institution</td>
<td>Role</td>
</tr>
<tr>
<td>------------------------------</td>
<td>----------------------------------</td>
<td>------------------------------------------------</td>
</tr>
<tr>
<td>Marinus J.C. Eijkemans, PhD</td>
<td>UMC Utrecht, Utrecht, The Netherlands</td>
<td>Designed and conceptualized study; Interpreted the data; revised the manuscript for intellectual content</td>
</tr>
<tr>
<td>Leonard H. van den Berg, MD, PhD</td>
<td>UMC Utrecht, Utrecht, The Netherlands</td>
<td>Designed and conceptualized study; Interpreted the data; revised the manuscript for intellectual content</td>
</tr>
</tbody>
</table>
REFERENCES


9. Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?". Lancet 2005;365:82-93.


**Figure 1.** Distribution of risk profiles in clinical trial and population-based cohorts

**Figure 1 legend.** For each patient we calculated his or her risk profile (horizontal axis), a composite score of seven 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 grey. 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 under-representation of patients with a poor prognosis in clinical trials. Exact ranges per cohort are available from Dryad (eAppendix 1; https://doi.org/10.5061/dryad.fbg79cnv7).
Figure 2. Expected trial duration under the null and alternative hypothesis.

Figure 2 legend. Distribution of simulated trial duration under the null hypothesis (H₀, i.e. HR = 1, panel A) and alternative hypothesis (H₁, i.e. HR = 0.63, panel B). Abbreviations: FPI = first-patient-first-visit; PI = percentile interval of the empirical trial duration ranging from the 2.5th to the 97.5th percentile; HR = hazard ratio. 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 grey is a scenario when the survival probability in the placebo arm is better than expected (60.0% instead of 56.1%).
Table 1. Overview of the interaction between disease progression rate and individual risk profile

<table>
<thead>
<tr>
<th>Clinical trial</th>
<th>Year</th>
<th>No. of patients (No. of observations)</th>
<th>Rate of progression (Points per month)</th>
<th>Interaction risk profile - progression rate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Points per month</td>
</tr>
<tr>
<td>Creatine</td>
<td>2001</td>
<td>175 (777)</td>
<td>-0.95</td>
<td>-0.22</td>
</tr>
<tr>
<td>Valproic acid</td>
<td>2009</td>
<td>163 (717)</td>
<td>-1.01</td>
<td>-0.14</td>
</tr>
<tr>
<td>LITRA</td>
<td>2012</td>
<td>133 (782)</td>
<td>-1.11</td>
<td>-0.18</td>
</tr>
<tr>
<td>LiCALS</td>
<td>2013</td>
<td>214 (1,125)</td>
<td>-1.05</td>
<td>-0.22</td>
</tr>
<tr>
<td>Ozanezumab</td>
<td>2017</td>
<td>303 (3,493)</td>
<td>-0.92</td>
<td>-0.23</td>
</tr>
<tr>
<td>PRO-ACT I</td>
<td>1996 - 2016</td>
<td>1,688 (17,942)</td>
<td>-1.04</td>
<td>-0.20</td>
</tr>
<tr>
<td>PRO-ACT II</td>
<td>1996 - 2016</td>
<td>748 (6,231)</td>
<td>-0.98</td>
<td>-0.19</td>
</tr>
<tr>
<td><strong>Pooled estimate:</strong></td>
<td></td>
<td></td>
<td><strong>-1.01</strong></td>
<td><strong>-0.20</strong></td>
</tr>
</tbody>
</table>

Rate of progression is based on the ALSFRS (creatine trial & PRO-ACT II) or ALSFRS-R. The PRO-ACT database was split into two 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. Abbreviations: CI = confidence interval. 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 Rubin’s rule. The pooled estimate is the meta-analyzed average across trial populations.
Table 2. Population-based characteristics of eligible patients at diagnosis

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>All patients (N = 5,100)</th>
<th>Eligible patients (N = 3,872)</th>
<th>Excluded patients</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(N = 793)</td>
<td>(N = 435)</td>
</tr>
<tr>
<td>Age at diagnosis, years</td>
<td>65 (11)</td>
<td>65 (11)</td>
<td>60 (14)</td>
</tr>
<tr>
<td>Bulbar site of onset</td>
<td>1,618 (32%)</td>
<td>1,270 (33%)</td>
<td>94 (12%)</td>
</tr>
<tr>
<td>Diagnostic delay, months*</td>
<td>10 (11)</td>
<td>10 (7)</td>
<td>29 (23)</td>
</tr>
<tr>
<td>FVC, %predicted</td>
<td>84 (22)</td>
<td>85 (21)</td>
<td>92 (19)</td>
</tr>
<tr>
<td>ALSFRS-R total score</td>
<td>38 (7)</td>
<td>38 (6)</td>
<td>40 (6)</td>
</tr>
<tr>
<td>ΔFRS, points per months*</td>
<td>-0.68 (1.03)</td>
<td>-0.74 (0.84)</td>
<td>-0.23 (0.16)</td>
</tr>
<tr>
<td>Definite El Escorial category</td>
<td>1,330 (26%)</td>
<td>961 (25%)</td>
<td>132 (17%)</td>
</tr>
<tr>
<td>Presence of ALS-FTD</td>
<td>307 (6.0%)</td>
<td>216 (5.6%)</td>
<td>25 (3.2%)</td>
</tr>
<tr>
<td>12-month survival since diagnosis</td>
<td>66.9%</td>
<td>67.3%</td>
<td>87.6%</td>
</tr>
<tr>
<td>Estimated median survival, months**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Receiving placebo</td>
<td>18.7</td>
<td>18.6</td>
<td>37.4</td>
</tr>
<tr>
<td>Receiving treatment (HR of 0.63)</td>
<td>27.4 (+8.7)</td>
<td>27.2 (+8.6)</td>
<td>52.0 (+14.6)</td>
</tr>
</tbody>
</table>
Patients < -6.0 are patients with an excellent prognosis, whereas patients > -2.0 are patients with a poor prognosis. Data are given in mean (SD) or % (n). *Data are median (interquartile range). **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. Abbreviations: FTD = frontotemporal dementia; ALSFRS-R = ALS functional rating scale; FVC = Forced Vital Capacity; ΔFRS = ALSFRS-R score - 48 / diagnostic delay; HR = hazard ratio.
Table 3. Classical vs. event-driven, GSD design of a hypothetical clinical trial for ALS

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Classical design</th>
<th>Event-driven, GSD design</th>
</tr>
</thead>
<tbody>
<tr>
<td>Planned sample size</td>
<td>400</td>
<td>278 (-30.5%)</td>
</tr>
<tr>
<td>Required number of events</td>
<td>148</td>
<td>153 (+2.9%)</td>
</tr>
<tr>
<td>Empirical power</td>
<td>79.5%</td>
<td>80.4%</td>
</tr>
<tr>
<td>Trial duration (months)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HR = 1.00 (H₀)</td>
<td>35.3</td>
<td>22.8 (-35.4%)</td>
</tr>
<tr>
<td>HR = 0.63 (H₁)</td>
<td>35.3</td>
<td>26.9 (-23.8%)</td>
</tr>
<tr>
<td>Total placebo exposure (person-months)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HR = 1.00 (H₀)</td>
<td>3,012</td>
<td>1,947 (-35.4%)</td>
</tr>
<tr>
<td>HR = 0.63 (H₁)</td>
<td>3,012</td>
<td>2,249 (-25.3%)</td>
</tr>
<tr>
<td>Average placebo exposure (months)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HR = 1.00 (H₀)</td>
<td>15.1</td>
<td>14.0 (-7.0%)</td>
</tr>
<tr>
<td>HR = 0.63 (H₁)</td>
<td>15.1</td>
<td>16.2 (+7.4%)</td>
</tr>
<tr>
<td>Follow-up costs (million $)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HR = 1.00 (H₀)</td>
<td>2.94</td>
<td>1.90 (-35.2%)</td>
</tr>
<tr>
<td>HR = 0.63 (H₁)</td>
<td>3.02</td>
<td>2.18 (-27.7%)</td>
</tr>
</tbody>
</table>

Abbreviations: HR = hazard ratio; GSD = group-sequential design. Classical design illustrates a 1:1 randomized trial with a fixed follow-up of 18-months and 80% power to detect a 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 is defined as the sum of individual follow-up costs for all patients (data are available from Dryad (eAppendix 1; https://doi.org/10.5061/dryad.fbg79cnv7).
Challenging the Established Order: Innovating Clinical Trials for Amyotrophic Lateral Sclerosis
Ruben P.A. van Eijk, Stavros Nikolakopoulos, Kit C.B. Roes, et al.
Neurology published online July 27, 2021
DOI 10.1212/WNL.0000000000012545

This information is current as of July 27, 2021

Updated Information & Services
including high resolution figures, can be found at:
http://n.neurology.org/content/early/2021/07/27/WNL.0000000000012545.full

Subspecialty Collections
This article, along with others on similar topics, appears in the following collection(s):
All Clinical trials
http://n.neurology.org/cgi/collection/all_clinical_trials
Amyotrophic lateral sclerosis
http://n.neurology.org/cgi/collection/amyotrophic_lateral_sclerosis
Clinical trials Methodology/study design
http://n.neurology.org/cgi/collection/clinical_trials_methodology_study_design

Permissions & Licensing
Information about reproducing this article in parts (figures, tables) or in its entirety can be found online at:
http://www.neurology.org/about/about_the_journal#permissions

Reprints
Information about ordering reprints can be found online:
http://n.neurology.org/subscribers/advertise

Neurology® is the official journal of the American Academy of Neurology. Published continuously since 1951, it is now a weekly with 48 issues per year. Copyright Copyright © 2021 The Author(s). Published by Wolters Kluwer Health, Inc. on behalf of the American Academy of Neurology. All rights reserved. Print ISSN: 0028-3878. Online ISSN: 1526-632X.