



OPEN ACCESS

ORIGINAL RESEARCH

Critical design considerations for time-to-event endpoints in amyotrophic lateral sclerosis clinical trials

Ruben P A van Eijk ^{1,2}, Stavros Nikolakopoulos,² Kit C B Roes,² Bas M Middelkoop,¹ Toby A Ferguson,³ Pamela J Shaw,⁴ P Nigel Leigh,⁵ Ammar Al-Chalabi,⁶ Marinus J C Eijkemans,² Leonard H van den Berg¹

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/jnnp-2019-320998>).

For numbered affiliations see end of article.

Correspondence to

Dr Ruben P A van Eijk, Department of Neurology, UMC Utrecht Brain Center Rudolf Magnus, Utrecht 3584 CG, The Netherlands; R.P.A.vanEijk-2@umcutrecht.nl

Received 14 April 2019

Revised 31 May 2019

Accepted 13 June 2019

Published Online First 10 July 2019

ABSTRACT

Background Funding and resources for low prevalent neurodegenerative disorders such as amyotrophic lateral sclerosis (ALS) are limited, and optimising their use is vital for efficient drug development. In this study, we review the design assumptions for pivotal ALS clinical trials with time-to-event endpoints and provide optimised settings for future trials.

Methods We extracted design settings from 13 completed placebo-controlled trials. Optimal assumptions were estimated using parametric survival models in individual participant data (n=4991). Designs were compared in terms of sample size, trial duration, drug use and costs.

Results Previous trials overestimated the hazard rate by 18.9% (95% CI 3.4% to 34.5%, p=0.021). The median expected HR was 0.56 (range 0.33–0.66). Additionally, we found evidence for an increasing mean hazard rate over time (Weibull shape parameter of 2.03, 95% CI 1.93 to 2.15, p<0.001), which affects the design and planning of future clinical trials. Incorporating accrual time and assuming an increasing hazard rate at the design stage reduced sample size by 33.2% (95% CI 27.9 to 39.4), trial duration by 17.4% (95% CI 11.6 to 23.3), drug use by 14.3% (95% CI 9.6 to 19.0) and follow-up costs by 21.2% (95% CI 15.6 to 26.8).

Conclusions Implementing distributional knowledge and incorporating accrual at the design stage could achieve large gains in the efficiency of ALS clinical trials with time-to-event endpoints. We provide an open-source platform that helps investigators to make more accurate sample size calculations and optimise the use of their available resources.

INTRODUCTION

The extensive clinical heterogeneity of patients with low prevalent neurodegenerative disorders, such as amyotrophic lateral sclerosis (ALS), complicates the design of clinical trials.^{1–3} Disease heterogeneity inflates the variability in endpoints and, subsequently, the number of patients to demonstrate therapeutic efficacy.^{1,3} In addition, large treatment effects in these disorders are unlikely, forcing investigators to search for subtle proofs of efficacy.⁴ Relatively large sample sizes are, therefore, needed in a rare disorder, which stresses the importance of optimising clinical trial design.

ALS is characterised by a progressive loss of motor neurons, leading to muscle weakness and, ultimately, death due to respiratory failure within 3–5 years.^{5,6} As the life expectancy of patients with ALS is significantly reduced, survival time is one of the preferred outcomes for assessing treatment efficacy in pivotal ALS clinical trials.^{7–9} Although survival time could be influenced by interventions such as gastrostomy or respiratory support and may not be optimal to detect early signals for efficacy,¹ showing treatment benefit in terms of an increased life expectancy leaves little doubt about a drug's therapeutic potential. In contrast, treatment benefits based solely on endpoints evaluating physical functioning (eg, Edaravone),¹⁰ could be deemed too marginal to warrant drug approval or to stop a trial during interim analyses.^{11–14} Combining survival time with other endpoints (eg, daily functioning, muscle strength or lung function) can considerably increase the efficiency of clinical trials.¹⁵ Measuring survival time in clinical trials for ALS is, therefore, indispensable and optimising trial design for time-to-events endpoints (ie, survival time) may improve the overall efficiency of pivotal ALS clinical trials.

Trial design for time-to-event endpoints, and in particular sample size calculation, is, however, complex and depends on an array of assumptions. For example, the number of events (eg, deaths) during follow-up depends, among other things, on the hazard rate, accrual rate, dropout rate and expected treatment effect.^{16–18} Each of these variables has its own inherent assumptions, and misspecification of a single variable could negatively affect statistical power or unnecessarily expose patients to inferior treatment regimes.¹⁷

It has proved difficult to define evidence-based guidance for ALS trial design, which drives current pivotal ALS trials to use arbitrary and simplified assumptions. In this study, we evaluate the basic components of ALS trial design for time-to-event endpoints and provide optimised settings for future trials. We then introduce an open-access, web-based platform that assists investigators to implement and standardise optimal clinical trial methodology for ALS; a framework that could easily be extended to other fields with time-to-event endpoints.



► <http://dx.doi.org/10.1136/jnnp-2019-321411>



© Author(s) (or their employer(s)) 2019. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

To cite: van Eijk RPA, Nikolakopoulos S, Roes KCB, et al. *J Neurol Neurosurg Psychiatry* 2019;**90**:1331–1337.

METHODS

A three-step approach was used to develop a framework for ALS clinical trials with time-to-event endpoints. First, we searched the literature for ALS clinical trials to compose a list of historical assumptions. We then obtained individual trial participant data to (1) obtain real-world estimates and (2) evaluate historical design settings. Finally, we validated designs and sample size calculations by means of simulation and incorporated the results in a web-based platform.

Literature search and data extraction

The systematic literature search and selection of studies has been described in more detail elsewhere.¹⁹ In short, all placebo-controlled trials evaluating the efficacy of a single pharmaceutical agent, published in the period 2000–2018 and having survival as (co)-primary endpoint, were enrolled in this study. From each trial, we extracted the sample size, accrual time (date last enrolled patient – date first enrolled patient), minimum follow-up period, predicted probability of survival in the placebo arm, hypothesised treatment effect (ie, HR), alpha (type I error), power (1 – type II error) and the definition of an event. A plot digitalizer was then used to extract the observed survival probability from the published Kaplan-Meier curves.

Individual participant data

The Pooled Resource Open-Access ALS Clinical Trials (PRO-ACT) database (version December 2015) was used as primary database to obtain real-world estimates of survival patterns among patients with ALS.²⁰ PRO-ACT contains data for 10 731 individuals from 23 trials performed over the past 20 years, is IRB-approved and uses solely anonymised data. Patients provided their consent to participate during the individual trials. For each PRO-ACT individual, we reconstructed their demographic, follow-up and survival data; if no survival data were available, patients were censored after their last known follow-up visit. We excluded participants without any follow-up time. The final dataset contained complete survival data from 3497 patients. In order to validate survival patterns, four external datasets that were not part of PRO-ACT were included from the following trials: Creatine (n=175),²¹ Valproic acid (n=163),²² LiCALS (n=214)²³ and EMPOWER (n=942).²⁴

Measures of effect

In order to evaluate how different design settings affect trial efficiency, we defined four measures of efficiency: (1) sample size, (2) trial duration, (3) drug usage and (4) follow-up costs. Trial duration was defined as the sum of the accrual period and the minimum follow-up duration (ie, date last follow-up – date enrolment first patient). Drug usage was defined as the sum of all follow-up time (in days) in the treatment arm. Costs of follow-up were calculated per patient based on the following hypothetical scenario: screening and start-up fee \$1500; monthly visits (eg, questionnaires and safety lab) \$250; bimonthly visits (eg, additional muscle and lung function testing) \$750 and biannual visits (eg, additional visit to neurologist) \$1000. As an example: a patient who is followed up for 10.3 months would have one screening session, two biannual visits (M0 and M6), four bimonthly visits (M2, M4, M8 and M10) and five monthly visits (M1, M3, M5, M7 and M9), resulting in a total costs of \$7750.

Statistical analysis

For all survival analyses, we used a two-sided log-rank test to compare the Kaplan-Meier curves of the placebo and active

treatment arms. The number of events drives the statistical power of the log-rank test and is, therefore, the primary estimate in sample size calculations. The number of events occurring during follow-up depends on a variety of variables; the primary determinants are the hazard rate, the total follow-up time and sample size. The hazard rate is a decisive design setting that heavily affects sample size calculations. In the case of a survival endpoint, where the event is death, the hazard rate can be interpreted as the instantaneous risk of dying.

In order to obtain insight into the underlying hazard rates in ALS clinical trials, we used parametric survival models following a Weibull distribution. The Weibull distribution allows time-varying hazard rates (eg, the risk of dying is different at the start from that at the end of the trial), whereas the exponential distribution assumes a constant hazard rate over time. The exponential distribution is a special case of the Weibull distribution with the shape parameter set to 1 (indicating no multiplicative effect of time on the hazard rate). Importantly, both distributions fulfil the proportional hazards assumption of the log-rank test.

Our first aim was to evaluate whether ALS survival patterns are best modelled by a constant (ie, exponential) or increasing (ie, Weibull) hazard rate. A likelihood ratio test was used to compare an exponential (Weibull shape 1) with the best fitting Weibull model (Weibull shape $\neq 1$). Models were fitted in each trial dataset individually. Subsequently, we pooled the log-transformed Weibull shape parameters using a fixed effects meta-analysis.¹⁹ In none of the trials there was an effect of treatment, we therefore did not distinguish between patients on active or placebo treatment. Nevertheless, we performed one sensitivity analysis using only the placebo arms of each cohort to evaluate the consistency of our results.

Our second aim was to evaluate how assuming a constant (Weibull shape=1) or increasing (Weibull shape $\neq 1$) hazard rate at the design stage affects sample size calculations. The required number of events to detect a given effect size (ie, HR) was estimated with the formulae provided by Schoenfeld and Freedman.^{25 26} The probability of an event for a constant hazard rate was calculated using the framework described in Abel *et al.*;¹⁷ we used a modified framework to allow accelerating hazard rates.²⁷ All sample size calculations assumed a 1:1 randomisation ratio, uniform accrual period (ie, constant recruitment rate) and no loss-to-follow-up other than death or administrative censoring. To validate sample size estimations, we simulated clinical trial survival data based on both the exponential and Weibull distributions, as described previously.^{14 15} All statistical analyses and simulations were programmed in the R package *Shiny* (V.1.1.0, Chang *et al.*, 2018); the source documentation can be found at <http://reactive.tricals.org>.

RESULTS

In total, we identified 14 trials with mortality as (co)-primary endpoint; one trial was excluded due to incomplete reporting (vitamin E, 2005).²⁸ Trials varied widely in predicted hazard rates, definitions of survival, follow-up durations and hypothesised treatment effects (table 1). The median expected treatment effect (ie, HR) was 0.56 (range 0.33–0.66). All trials assumed a constant (ie, exponential) hazard rate, with a mean of 3.0 (range 0.9–5.1) events per 100 person-months and a hypothesised survival probability in the placebo arm of 58% (range 40%–85%) after 18 months. On average, the survival probability was underestimated with an absolute difference of 5.8% (95% CI 1.0% to 10.6%), resulting in a mean overestimation of the hazard rate of 18.9% (95% CI 3.4% to 34.5%, $p=0.021$). Five trials (38.5%)

Table 1 Design characteristics and assumptions of placebo-controlled trials in ALS with mortality endpoints

Design characteristics					Assumed survival		Observed survival	
Treatment (year)	No. patients	Event	Follow-up	Treatment effect (HR)	Survival (%) (at month)	Hazard rate	Survival (%)	Hazard rate
Creatine (2003)	175	D, T, N	16	0.44	60 (16)	3.2	53.5	3.9
Xaliproden I (2004)	867	D, T, N	18	0.62	50 (18)	3.9	53.8	3.4
Xaliproden II (2004)	1210	D, T, N	18	0.66	57 (18)	3.1	62.3	2.6
Pentoxifylline (2006)	400	D	18	0.65	40 (18)	5.1	59.7	2.9
Valproic acid (2009)	163	D, T, N	16	0.56	60 (16)	3.2	75.1	1.8
Lithium (2012)	133	D, T, N	16	0.56	60 (16)	3.2	66.5	2.5
Pioglitazone (2012)	219	D	≥18	0.49	60 (18)	2.8	69.6	2.0
Dexpramipexole (2013)	942	D, T, N	12–18	0.63	80 (12)	1.9	80.7	1.8
Lithium (2013)	214	D	18	0.45	65 (18)	2.4	59.4	2.9
Ceftriaxone (2014)	513	D, T, N	≥12	0.66	75 (12)	2.4	74.6	2.4
Olesoxime (2014)	512	D, T, N	18	0.63	62 (18)	2.7	67.2	2.2
Erythropoietin (2015)	208	D, T, N	12	0.33	58 (12)	4.5	74.1	2.5
Ozanezumab (2017)	303	D	11	0.49	90 (11)	0.9	96.1	0.4

Follow-up time is given in months. Number of patients is for the total sample size. Hazard rates are given in number of events per 100 person-months. ALS, amyotrophic lateral sclerosis; D, death; N, non-invasive ventilation; T, tracheostomy.

overestimated the hazard rate by more than 25%. Only two trials (Ceftriaxone 2014 and Pioglitazone 2012) considered the accrual period when determining the sample size.^{29 30}

Hazard rates in ALS clinical trials

Using clinical trial data from five independent cohorts (n=4991), we evaluated the assumption of a constant hazard rate in ALS (figure 1). In all cohorts, we found evidence for an increasing mean hazard rate over time (pooled Weibull shape parameter of 2.03, 95% CI 1.93 to 2.15, p<0.001, Cochran's Q test p=0.12). This result was similar when analysing only patients allocated to placebo (2.22, 95% CI 2.05 to 2.40, p<0.001). For example, the mean hazard rate in PRO-ACT during month 0–3 was 0.3 deaths per 100 person-months, which increased to 2.7 deaths per 100 person-months during month 15–18. This finding has important consequences for the number of events: in a hypothetical trial of 100 patients, with an 80% survival after 12 months, a constant rate predicts 11, 20 and 28 deaths after 6, 12 and 18 months, whereas an increasing hazard rate (Weibull shape parameter of 2) predicts 5, 20 and 39 deaths, respectively. Although the number of events at 12 months is identical, before and after this time point, the constant hazard assumption severely overestimates and underestimates the event probability.

Spending time wisely: importance of accrual

The majority of the ALS trials followed patients for a fixed time period. However, as patients are not enrolled simultaneously, the actual trial duration (from start to last follow-up) is the follow-up duration plus the accrual (ie, enrolment) period. We illustrate this process for 15 EMPOWER patients in figure 2A. An option might be to extend follow-up for those patients enrolled first, until the last patient has completed follow-up (ie, variable follow-up, figure 2B). Consequently, patients remain longer in the trial, generate more events and, consequently, the required sample size could be reduced. Importantly, the total trial duration is identical in both settings. In figure 2C, we show that assuming a constant hazard rate, using the observed 12 month survival in EMPOWER, severely overestimates survival for this extended follow-up period and underestimates the total number of events during the entire trial (figure 2D).

Effects on historical trial designs

Finally, we evaluated how increasing hazards and accrual affect ALS trials in terms of sample size, trial duration, costs and drug exposure time. We redesigned each trial from table 1 according to the observed accrual period and its original hypotheses under (1) constant and increasing hazard rates and (2) with and without accrual; power, alpha and the treatment effect were fixed. The results are given in figure 3. When accrual is not incorporated, both the constant and increasing rate provide identical sample sizes. However, when assuming the true survival pattern is accelerating (figure 1), the constant hazard assumption underestimates drug product usage by up to 11.1% (mean 4.7%; 95% CI 3.1 to 6.3, p<0.001) and follow-up costs by up to 10.5% (mean 4.8%; 95% CI 3.3 to 6.3, p<0.001). Incorporating accrual time and assuming an increasing hazard rate at the design stage reduced sample size, on average, by 33.2% (95% CI 27.9 to 39.4, p<0.001), trial duration by 17.4% (95% CI 11.6 to 23.3, p<0.001), drug use by 14.3% (95% CI 9.6 to 19.0, p<0.001) and follow-up costs by 21.2% (95% CI 15.6 to 26.8, p<0.001) compared with the classical design (constant hazard rate without accrual).

Implementation tool for future trials

In order to improve the implementation of increasing hazard rates and accrual into trial designs for future trials, we developed an open-access platform for time-to-event sample size calculations based on the presented methodology (TRICALS-*Reactive*; <http://reactive.tricals.org>). The platform provides estimates for survival patterns, acceleration rates and sample sizes. Estimates are based on the PRO-ACT database. Online supplementary eTable 1 provides a comparison of the performance of the platform as compared with classical methods (modified to allow Weibull distributions).²⁷ As described previously,¹⁷ Freedman is slightly conservative, whereas Schoenfeld is more liberal. The platform performs similarly to Schoenfeld and closely matches its target power.

DISCUSSION

In this study, we evaluated key assumptions underlying time-to-event designs for ALS clinical trials. We found consistent evidence

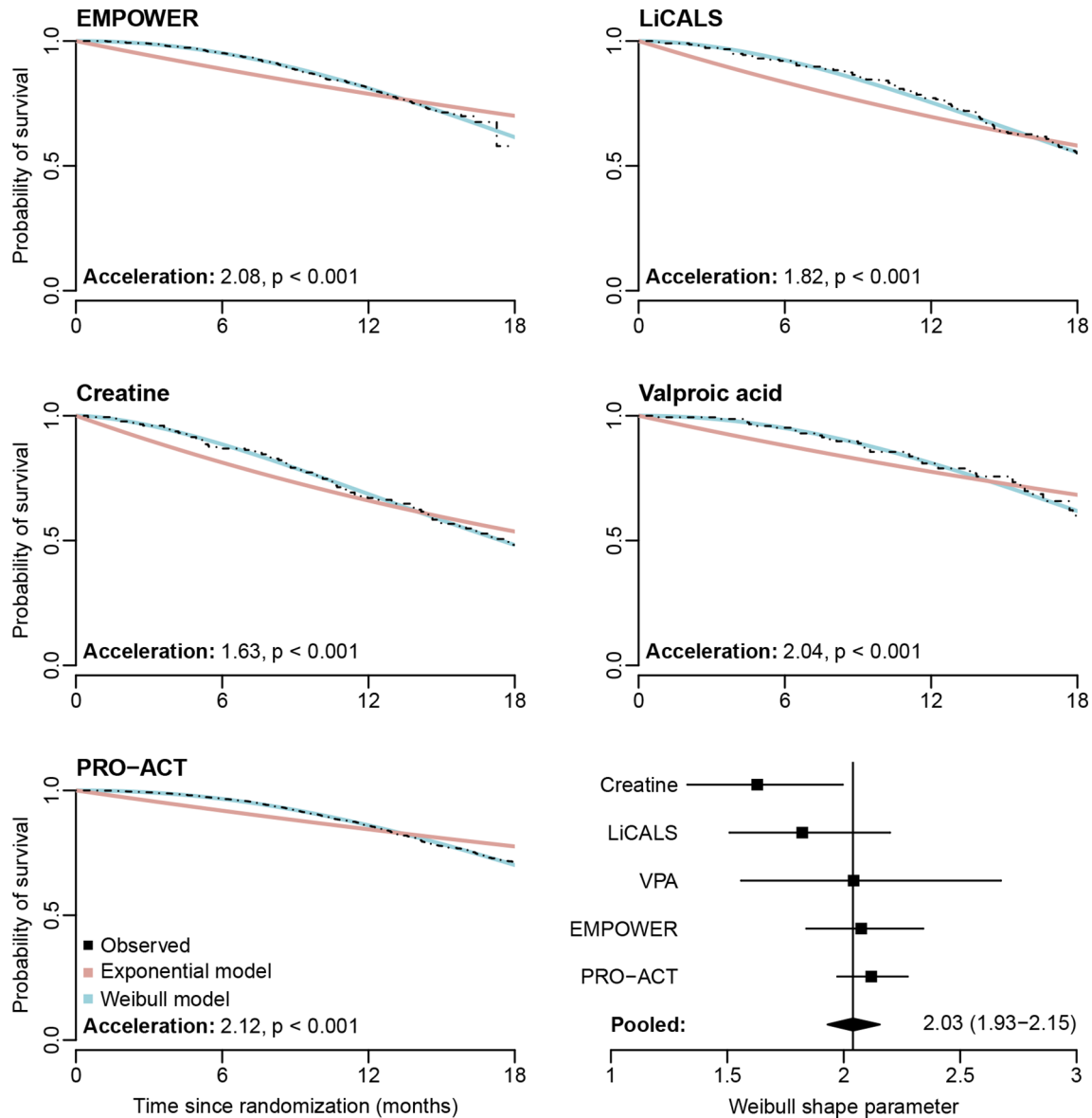


Figure 1 Constant versus increasing hazard rates in ALS clinical trials. Two models were fitted with either an exponential or Weibull distribution. The exponential model assumes a constant hazard rate over time (or Weibull shape parameter p of 1). Within each cohort, we determined whether the Weibull shape parameter p was different from 1. Results across cohorts are pooled by a fixed effects meta-analysis (lower right panel). ALS, amyotrophic lateral sclerosis; PRO-ACT, Pooled Resource Open-Access ALS Clinical Trials; VPA, Valproic acid study; EMPOWER, acronym of the dexamipexole study; LiCALS, acronym of the United Kingdom Lithium study.

for increasing hazard rates over time, which has important consequences for the design and planning of future clinical trials. Additionally, our results reveal a clear efficiency gain when the accrual period is incorporated into the design: extending the follow-up period for participants enrolled early in the trial considerably reduces sample sizes and, consequently, costs. As funds and resources for rare neurological disorders are limited, optimising their use is vital for efficient drug development. In order to mediate the implementation of our results, we developed an open-source, data-driven online platform that reduces the risk of mis-specification and facilitates sample size calculations for increasing hazards. The platform could be extended and applied in other disorders with time-to-event endpoints.

The need for sample size calculations is based on the ethical imperative to minimise the number of patients exposed to harmful experiments or inferior treatments,³¹ yet the number of enrolled patients needs to be sufficient to avoid falling short of

demonstrating efficacy, or worse, generating false conclusions. Sample size calculations are, therefore, a balancing act, where inaccurate estimation can unnecessarily put patients at risk or lead to a large waste of resources. Sample size calculations require the investigator to make several assumptions, which are well known to be arbitrary with a high risk of mis-specification.^{32 33} Our results reveal a systematic mis-specification among ALS clinical trials with an 18.9% overestimation of the hazard rate. Moreover, we found that, despite all trials being based on the same primary endpoint, trials varied widely in their definition of the event, effect size, follow-up duration and assumed survival probabilities. This variability affects sample size estimation and underscores the current lack of guidance in trial design for ALS.¹⁹ Improving the standardisation of trial design could improve the interpretation of trial results, facilitate cross-trial comparisons and potentially harmonise the reporting of trial results.³⁴ Our developed framework provides, along with data-driven

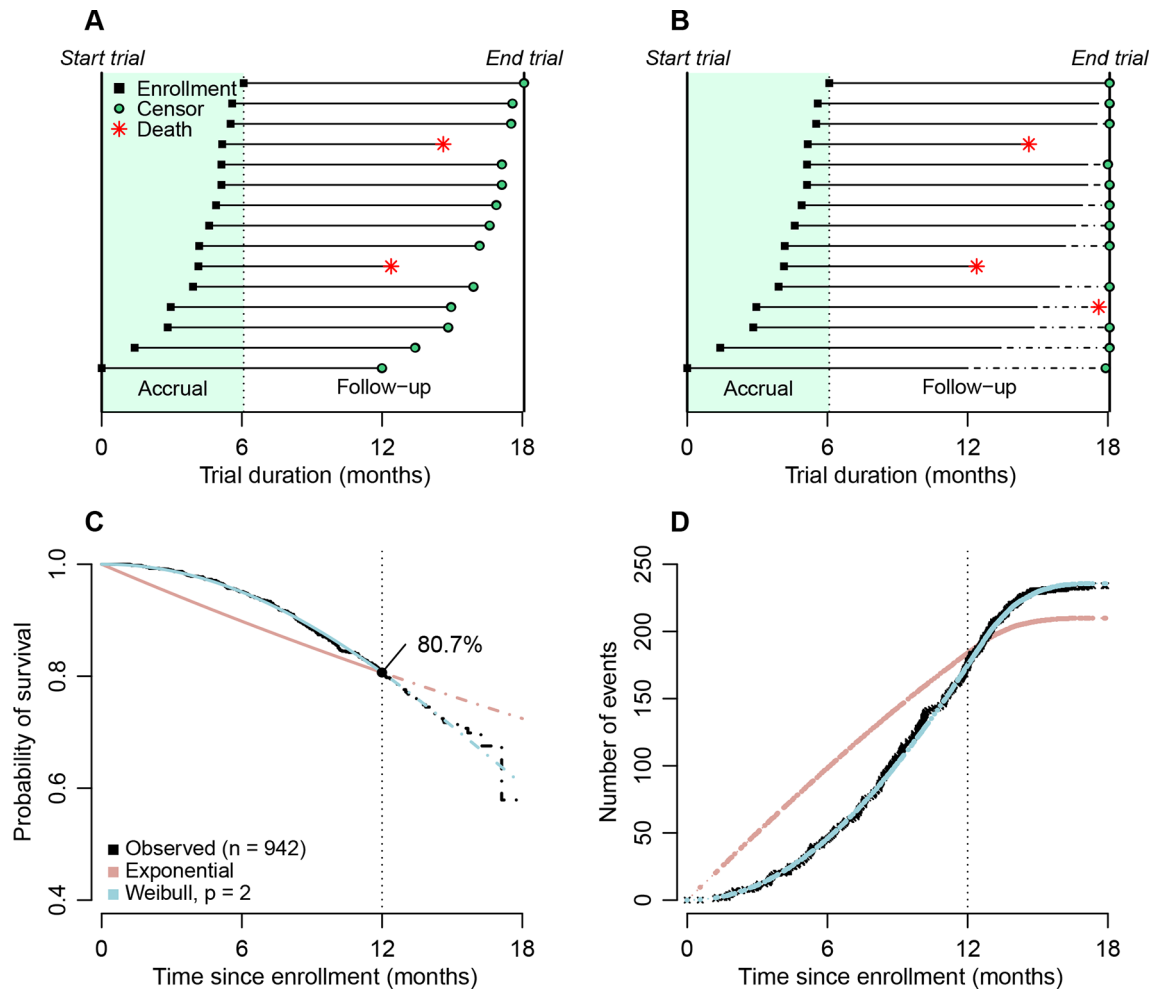


Figure 2 Trial duration, accrual and number of events in EMPOWER. (A) Classical trial design with fixed follow-up (here 12 months) for 15 EMPOWER patients. As patients are not all recruited at the same time, the total trial duration is the sum of the follow-up and accrual periods. (B) Extending follow-up until the last enrolled patient completed the 12 month follow-up increases the number of events and increases power. (C) Using the observed 12-month survival in EMPOWER, a constant (exponential) assumption underestimates survival before 12 months and subsequently overestimates survival. (D) Based on the EMPOWER data (n=942), we determined for each time point the expected number of events under the exponential and Weibull models. The black crosses are the observed events over time.

assumptions, a platform for the development and implementation of future guidelines for ALS clinical trials.

The systematic mis-specification in ALS clinical trials is partially the result of erroneously assuming a constant hazard rate during follow-up. Trial participants are a selected subpopulation,^{19,35} from which far-progressed patients in the end stage of ALS are excluded. Consequently, the event probability during a trial is initially low and increases over time. This acceleration factor was confirmed in our meta-analysis (figure 1). This has important consequences for sample size calculations and cost estimation and also for the planning of milestones. The Pioglitazone trial, for example, planned one interim analysis after 50% of the required number of events. Assuming a constant hazard, the interim analysis was planned after 12.3 months. After 12 months, however, the authors report that the number of events was *unexpectedly* low (~38% of required) and the interim analysis was abandoned.³⁰ If the trial had assumed an increasing hazard (eg, Weibull shape 2), the interim analysis would have been conducted 4.5 months later at 16.7 months.¹⁴ At that time, the authors would probably have reached their 50% target and might have stopped the trial earlier for reasons of futility, potentially reducing the loss of resources.

The efficiency of ALS trials with time-to-event endpoints could be further improved by including accrual at the design stage. However, early-enrolled participants would remain in the trial for a considerable amount of time, which may increase dropout rates. The feasibility of a prolonged follow-up period in ALS has been shown before by the Ceftriaxone trial,²⁹ and recently by a trial with methylcobalamin,³⁶ where some patients participated for over 4 years. Registry data from large longitudinal population-based cohorts could additionally provide insight into follow-up patterns, as patients are often monitored from diagnosis until death. Moreover, extending follow-up could improve a trial's generalisability to long-term treatments that patients with ALS will probably require. Nevertheless, incorporating accrual at the design stage does require accurate prior assumptions and this further underscores the necessity of (online) tools such as those presented here. Our study did not evaluate the accrual and dropout patterns in ALS trials, and incorporating separate models for accrual and dropout rates may further optimise their design and planning. As an alternative, a trial may continue to run until the required number of events is reached. This would make the trial duration (from first enrolment to last

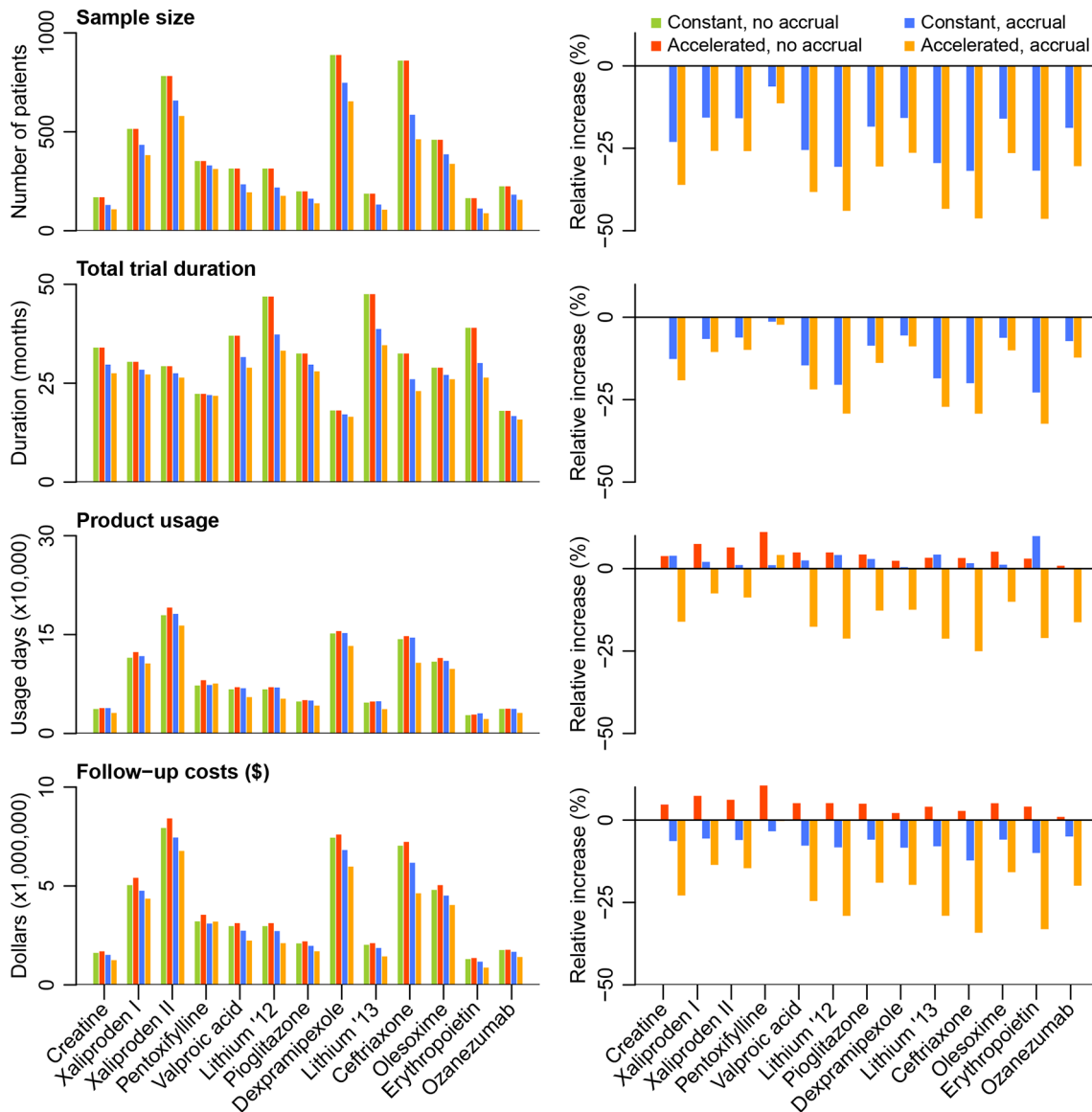


Figure 3 Effect of accrual and increasing hazards on historical trial designs. For each trial specified in table 1, we re-estimated the sample size according to the observed accrual period and the original design assumptions (x-axis). We provide in the left column barcharts the estimated sample size, total trial duration, product usage and follow-up costs and in the right column barcharts their respective relative increases as compared with the classical trial design (ie, constant hazard rate, without incorporating accrual). We used the formula of Schoenfeld to determine the number of events. Trial designs for accelerated (ie, increasing) hazards were based on a Weibull shape of 2.

follow-up visit) uncertain, but prevent losses in efficiency due to inaccurate assumptions of accrual, dropout and/or hazard rates.

As a final note, survival patterns in clinical trials may be affected by eligibility criteria, or geographical differences in care or infrastructure, which may affect sample size.¹⁹ To illustrate, if we use the default settings in our tool, the estimated sample size is 412 patients. When selecting patients between 60 and 70 years old, symptom duration of 12–18 months and %predicted lung function of 60%–90%, the sample size is reduced to 258. In this case, the difference in sample size is primarily driven by a difference in death rates (ie, 18 month survival of 71.2% vs 49.7%, respectively). In addition, survival patterns in future trials may alter due to improved diagnostic strategies or the use of combination therapies. Although our current tool has the ability to study the effect of eligibility criteria on sample size, more detailed data (eg, genotype or prediction-based information) may be required to accommodate all future settings. Moreover, the development

of similar modelling techniques for other efficacy endpoints such as the ALS functional rating scale or the joint modelling framework may further optimise the design of clinical trials.³⁷

In conclusion, resources for rare incurable disorders such as ALS are limited, and optimising their use is vital to increase the efficiency of drug development. Large reductions in sample size, duration and costs of ALS clinical trials with time-to-event endpoints could be achieved by implementing parametrical models that incorporate prior knowledge of the survival patterns and incorporating accrual at the design stage. We provide an open-source platform that helps investigators make more accurate sample size calculations and optimise the use of their available resources.

Author affiliations

¹Department of Neurology, Brain Center Rudolf Magnus, University Medical Centre Utrecht, Utrecht, The Netherlands

²Biostatistics & Research Support, Julius Centre for Health Sciences and Primary Care, Utrecht, The Netherlands

³Department of Neurology Research and Early Clinical Development, Biogen Inc, Cambridge, Massachusetts, USA

⁴Sheffield Institute for Translational Neuroscience, University of Sheffield, Sheffield, UK

⁵Department of Clinical Neuroscience, Trafford Centre for Biomedical Research, Brighton and Sussex Medical School, Brighton, UK

⁶Department of Basic and Clinical Neuroscience, Maurice Wohl Clinical Neuroscience Institute, King's College London, London, UK

Contributors RPAVE, SN: study concept, design, analysis, interpretation of data and drafting manuscript. KCBR, MJCE: study concept, design, analysis and interpretation of data. BMM: design and critical revision of manuscript for intellectual content. TAF, PJS, PNL, AA-C: acquisition of data, critical revision of manuscript for intellectual content. LHVB: study supervision and critical revision of manuscript for intellectual content.

Funding This study was funded by the Netherlands ALS Foundation (Project TRICALS-Reactive).

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available in a public, open access repository. All data relevant to the study are included in the article or uploaded as supplementary information.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

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

REFERENCES

- Berry JD, Cudkovic ME. New considerations in the design of clinical trials for amyotrophic lateral sclerosis. *Clin Investig* 2011;1:1375–89.
- 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;13:1127–38.
- Chio A, Canosa A, Gallo S, et al. ALS clinical trials: do enrolled patients accurately represent the ALS population? *Neurology* 2011;77:1432–7.
- Miller RG, Moore DH. ALS trial design: expectation and reality. *Amyotroph Lateral Scler Other Motor Neuron Disord* 2004;5 Suppl 1:52–4.
- van Es MA, Hardiman O, Chio A, et al. Amyotrophic lateral sclerosis. *The Lancet* 2017;390:2084–98.
- Knibb JA, Keren N, Kulka A, et al. A clinical tool for predicting survival in ALS. *J Neurol Neurosurg Psychiatry* 2016;87:1361–7.
- Shefner JM. Designing clinical trials in amyotrophic lateral sclerosis. *Phys Med Rehabil Clin N Am* 2008;19:495–508.
- Food Drug Administration Center for Drugs Evaluation Research. Guidance for Industry: Amyotrophic Lateral Sclerosis: Developing Drugs for Treatment, 2018. Available: <https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM596718.pdf> [Accessed 1 May 2019].
- 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 1 May 2019].
- Edaravone Writing Group. Safety and efficacy of edaravone in well defined patients with amyotrophic lateral sclerosis: a randomised, double-blind, placebo-controlled trial. *Lancet Neurol* 2017;16:505–12.
- van Eekelen R, de Hoop E, van der Tweel I. Simultaneous sequential monitoring of efficacy and safety led to masking of effects. *J Clin Epidemiol* 2016;76:155–65.
- Hardiman O, van den Berg LH. Edaravone: a new treatment for ALS on the horizon? *Lancet Neurol* 2017;16:490–1.
- Al-Chalabi A, Andersen PM, Chandran S, et al. ENCALS statement on edaravone. *Amyotroph Lateral Scler Frontotemporal Degener* 2017;2017:471–4.
- van Eijk RPA, Nikolakopoulos S, Ferguson TA, et al. Increasing the efficiency of clinical trials in neurodegenerative disorders using group sequential trial designs. *J Clin Epidemiol* 2018;98:80–8.
- van Eijk RP, Eijkemans MJ, Rizopoulos D, et al. Comparing methods to combine functional loss and mortality in clinical trials for amyotrophic lateral sclerosis. *Clin Epidemiol* 2018;10:333–41.
- Schoenfeld DA. Sample-size formula for the proportional-hazards regression model. *Biometrics* 1983;39:499–503.
- Abel UR, Jensen K, Karapanagiotou-Schenkel I, et al. Some issues of sample size calculation for Time-to-Event endpoints using the Freedman and Schoenfeld formulas. *J Biopharm Stat* 2015;25:1285–311.
- Lachin JM, Foulkes MA. Evaluation of sample size and power for analyses of survival with allowance for nonuniform patient entry, losses to follow-up, noncompliance, and stratification. *Biometrics* 1986;42:507–19.
- van Eijk RPA, Westeneng H-J, Nikolakopoulos S, et al. Refining eligibility criteria for amyotrophic lateral sclerosis clinical trials. *Neurology* 2019;92:e451–60.
- Atassi N, Berry J, Shui A, et al. The PRO-ACT database: design, initial analyses, and predictive features. *Neurology* 2014;83:1719–25.
- Groeneweld GJ, Veldink JH, van der Tweel I, et al. A randomized sequential trial of creatine in amyotrophic lateral sclerosis. *Ann Neurol* 2003;53:437–45.
- Piepers S, Veldink JH, de Jong SW, et al. Randomized sequential trial of valproic acid in amyotrophic lateral sclerosis. *Ann Neurol* 2009;66:227–34.
- 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;12:339–45.
- Cudkovic ME, van den Berg LH, Shefner JM, et al. Dexamipexole versus placebo for patients with amyotrophic lateral sclerosis (empower): a randomised, double-blind, phase 3 trial. *Lancet Neurol* 2013;12:1059–67.
- Freedman LS. Tables of the number of patients required in clinical trials using the logrank test. *Stat Med* 1982;1:121–9.
- Schoenfeld DA, Richter JR. Nomograms for calculating the number of patients needed for a clinical trial with survival as an endpoint. *Biometrics* 1982;38:163–70.
- Wu JR. Power and sample size for randomized phase III survival trials under the Weibull model. *J Biopharm Stat* 2015;25:16–28.
- Graf M, Ecker D, Horowski R, et al. High dose vitamin E therapy in amyotrophic lateral sclerosis as add-on therapy to riluzole: results of a placebo-controlled double-blind study. *J Neural Transm* 2005;112:649–60.
- Cudkovic ME, Titus S, Kearney M, et al. Safety and efficacy of ceftriaxone for amyotrophic lateral sclerosis: a multi-stage, randomised, double-blind, placebo-controlled trial. *Lancet Neurol* 2014;13:1083–91.
- Dupuis L, Dengler R, Heneka MT, et al. A randomized, double blind, placebo-controlled trial of pioglitazone in combination with riluzole in amyotrophic lateral sclerosis. *PLoS One* 2012;7:e37885.
- Cesana BM, Antonelli P. Sample size calculations in clinical research should also be based on ethical principles. *Trials* 2016;17.
- Nikolakopoulos S, Roes KCB, van der Lee JH, et al. Sample size calculations in pediatric clinical trials conducted in an ICU: a systematic review. *Trials* 2014;15.
- 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:e0132578.
- Gourgou-Bourgade S, Cameron D, Poortmans P, et al. Guidelines for time-to-event end point definitions in breast cancer trials: results of the DATECAN initiative (Definition for the Assessment of Time-to-event Endpoints in CANcer trials)†. *Ann Oncol* 2015;26:873–9.
- 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;88:557–63.
- Kaji R, Imai T, Iwasaki Y, et al. Ultra-high-dose methylcobalamin in amyotrophic lateral sclerosis: a long-term phase II/III randomised controlled study. *J Neurol Neurosurg Psychiatry* 2019;90:451–7.
- 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.