Hybrid Controlled Clinical Trials Using Concurrent Registries in Amyotrophic Lateral Sclerosis: A Feasibility Study

Ruben P. A. van Eijk[1,2,*], Leonard H. van den Berg[2], Kit C. B. Roes[3], Lu Tian[1], Tze L. Lai[1], Lorene M. Nelson[4], Chenyu Li[1], Anna Scowcroft[5], Jesus Garcia-Segovia[5] and Ying Lu[1,4,*]

Hybrid designs with both randomized arms and an external control cohort preserve key features of randomization and utilize external information to augment clinical trials. In this study, we propose to leverage high-quality, patient-level concurrent registries to enhance clinical trials and illustrate the impact on trial design for amyotrophic lateral sclerosis. The proposed methodology was evaluated in a randomized, placebo-controlled clinical trial. We used patient-level information from a well-defined, population-based registry, that was running parallel to the randomized clinical trial, to identify concurrently nonparticipating, eligible patients who could be matched with trial participants, and integrate them into the statistical analysis. We assessed the impact of the addition of the external controls on the treatment effect estimate, precision, and time to reach a conclusion. During the runtime of the trial, a total of 1,141 registry patients were alive; 473 (41.5%) of them fulfilled the eligibility criteria and 133 (11.7%) were enrolled in the study. A matched control population could be identified among the nonparticipating patients. Augmenting the randomized controls with matched external controls could have avoided unnecessary randomization of 17 patients (−12.8%) as well as reducing the study duration from 30.1 months to 22.6 months (−25.0%). Matching eligible external controls from a different calendar period led to bias in the treatment effect estimate. Hybrid trial designs utilizing a concurrent registry with rigorous matching can minimize bias due to a mismatch in calendar time and differences in standard of care, and may accelerate the development of new treatments.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?

☑ Novel trial methodology is needed to combine randomized clinical trials and real-world data. Previous studies have used historical controls to augment randomized controls; such studies are, however, at risk of bias and produce a lower level of evidence.

WHAT QUESTION DID THIS STUDY ADDRESS?

☑ We aimed to identify external cohorts that are interchangeable with randomized control arms by using high-quality, patient-level concurrent registries that run in parallel with randomized clinical trials.

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?

☑ We illustrate how a concurrent registry can be leveraged to integrate external controls into a hybrid design for amyotrophic lateral sclerosis, thereby improving precision, reducing the time to reach a decision, and lowering the number of patients randomized and allocated to placebo.

HOW MIGHT THIS CHANGE CLINICAL PHARMACOLOGY OR TRANSLATIONAL SCIENCE?

☑ Especially for diseases with a rare and significant unmet medical need, with limited treatment options available, a hybrid design of clinical trials allows studies to be conducted when large placebo arms are unethical or infeasible, accelerating the development of new treatments.

[1]Department of Biomedical Data Science and Centre for Innovative Study Design, School of Medicine, Stanford University, Stanford, California, USA; [2]Department of Neurology, UMC Utrecht Brain Centre, University Medical Centre Utrecht, Utrecht, The Netherlands; [3]Section Biostatistics, Department of Health Evidence, Radboud Medical Centre, Nijmegen, The Netherlands; [4]Department of Epidemiology and Population Health, School of Medicine, Stanford University, Stanford, California, USA; [5]UCB Pharma, Brussels, Belgium. *Correspondence: Ruben P.A. van Eijk (r.p.a.vaneijk-2@umcutrecht.nl) and Ying Lu (ylu1@stanford.edu)

Received February 3, 2023; accepted June 27, 2023. doi:10.1002/cpt.2994

Randomized clinical trials (RCTs) have been foundational for demonstrating drug efficacy and regulatory decision making. The costs of clinical trials, however, have been rising exponentially and are a major driver behind pharmaceutical prices.[1] Especially in rare diseases, the success rate of clinical trials remains poor due to low disease prevalence, lack of sensitive end points, and often considerable phenotypic heterogeneity.[2] Major regulatory agencies in the United States and Europe recognize the need for alternative approaches, thereby actively encouraging the development of innovative trial designs.[3,4] The US Food and Drug Administration (FDA) has launched two programs, the Complex Innovative Trial Designs Pilot Program[5] and Clinical Trials Transformation Initiative,[6] and the European Medicines Agency initiated the Accelerating Clinical Trials in the European Union (ACT EU),[7] which encourages the use of real-world data (RWD) to supplement clinical trials.

One of these approaches is the use of RWD, originating from registries, electronic health records, and claims databases to augment clinical trials and support decision making regarding the benefits and risks of a medicinal product.[8–11] Of particular interest is the use of RWD as external control to replace or reduce the size of the randomized control arm in RCTs. Specifically for diseases with a significant unmet medical need, external controls can enable studies to be conducted when complete randomization is unethical or infeasible, can replace the need for a placebo control and/or substantially reduce sample size requirements. This is not without risk, however, as the integration of external data, with patient-level information, could distort the trial's integrity and may introduce substantial bias.[12]

As a result, there is a need for the development of novel trial methodologies capable of combining the best aspects of RCTs and RWD. Hybrid designs with both a randomized control arm and external control data preserve key features of RCTs, such as randomization, and benefit from RWD by augmenting, instead of replacing, the randomized controls.[13] Significantly, these designs allow for a fallback mechanism if the randomized and external controls differ substantially, by having the option to solely utilize the randomized controls for the final analysis,[14] and thereby protect the integrity of the treatment comparison.

The value of hybrid designs, therefore, depends on the ability to identify an external cohort that is interchangeable with the randomized control arm.[15] In this study, we approach this challenge by proposing the use of a high-quality, patient-level concurrent registry that runs in parallel with the RCT. Important sources of bias, such as a mismatch in calendar time and geographic differences in standard of care, can thereby be minimized.[16] We illustrate the feasibility and validity of this approach in amyotrophic lateral sclerosis (ALS) where, historically, 20 to 25% of the eligible patients participate in clinical trials.[17] The registry is used to identify nonparticipating eligible patients, and concurrently match these nonparticipants with trial participants to minimize potential differences in rates of progression and survival among control groups. Finally, we augment the control arm and make a more efficient inference about the treatment effect.

METHODS

The original clinical trial and registry study were reviewed and approved by the Institutional Review Board (IRB) of the University Medical Center Utrecht (UMC, Utrecht, The Netherlands), and conducted according to the Medical Research Involving Human Subjects Act (WMO) and the International Conference on Harmonization Good Clinical Practice guidelines. All patients in the clinical trial and registry provided written informed consent. Re-use of the data was exempted from review by the IRBs of the UMC Utrecht and Stanford University (CA, USA) as no new data were collected, nor shared with third parties outside of the UMC Utrecht.

Clinical trial

A completed, randomized, placebo-controlled clinical trial was used to illustrate the feasibility of the proposed methodology (trial registration number NTR1448).[18] The trial aimed to determine the safety and efficacy of lithium carbonate. The rationale for the trial was based on a relatively small pilot study that found a significant effect on survival along with a slowing of disease progression.[19] Patients with a diagnosis of ALS, according to the World Federation of Neurology El Escorial criteria,[20] were enrolled at three national referral centers in the Netherlands, and 1:1 randomized between November 2008 and June 2011. Patients received either lithium carbonate at a target concentration of 0.4–0.8 mEq/L or a matching placebo. Other major inclusion criteria required an onset of symptoms at least 6 months and no longer than 36 months prior to inclusion, and a sitting forced vital capacity (FVC) of at least 70% of the predicted value based on gender, height, and age. Patients in the trial were treated for up to 30 months. The primary end point was time to a composite end point, defined as the time from inclusion to death, tracheostomy or noninvasive ventilation for more than 16 hours per day. A sequential design was used to detect a hazard ratio (HR) of 0.56 with 90% power and a one-sided alpha of 5%; the expected sample size was 173 patients if treatment were futile and 191 patients if the alternative hypothesis were true. The trial was stopped for futility when 61 of the 133 patients reached the primary end point (66 allocated to lithium and 67 allocated to placebo). The maximum theoretical follow-up time in the trial (i.e., from start of enrollment to last follow-up), was 30.9 months.

Population-based registry

We sought concurrent controls from the same source population that was used to recruit patients for the clinical trial.[21] The Netherlands ALS Registry is a prospective, population-based registry since April 2006, which was used to identify patients who were not enrolled in the clinical trial. Patients are identified by annual screening of large medical center registries and by individually contacting Dutch neurologists. The registry collects patient characteristics from the day of diagnosis. Complete mortality data are obtained by examining the online municipal population register at quarterly intervals. To harmonize the outcome data between the trial and the registry,[22] we redefined the trial primary end point as time to death only (instead of the composite end point), thereby updating the survival data of the clinical trial and re-analyzing the original study (**Figure S1**).

Identifying concurrent controls

The selection and identification of suitable concurrent external control patients in the registry is illustrated in **Figure 1**. The runtime of the trial (i.e., from the start of enrollment to the last follow-up), is depicted in green. To identify all patients who could have participated in the trial (i.e., the eligible population), we selected those patients who were either diagnosed before or during the trial enrollment period and who were alive at the start of the study. As such, all patients in the registry who died before the start of the trial, or those who were diagnosed after the trial had been completed, were excluded as potential candidates (illustrated as patients 1–4). Subsequently, we applied the trial eligibility criteria to the remaining patients. As key inclusion criteria after diagnosis were not collected systematically, trial eligibility was assessed as on the day of diagnosis. This timepoint may be earlier than a patient would normally

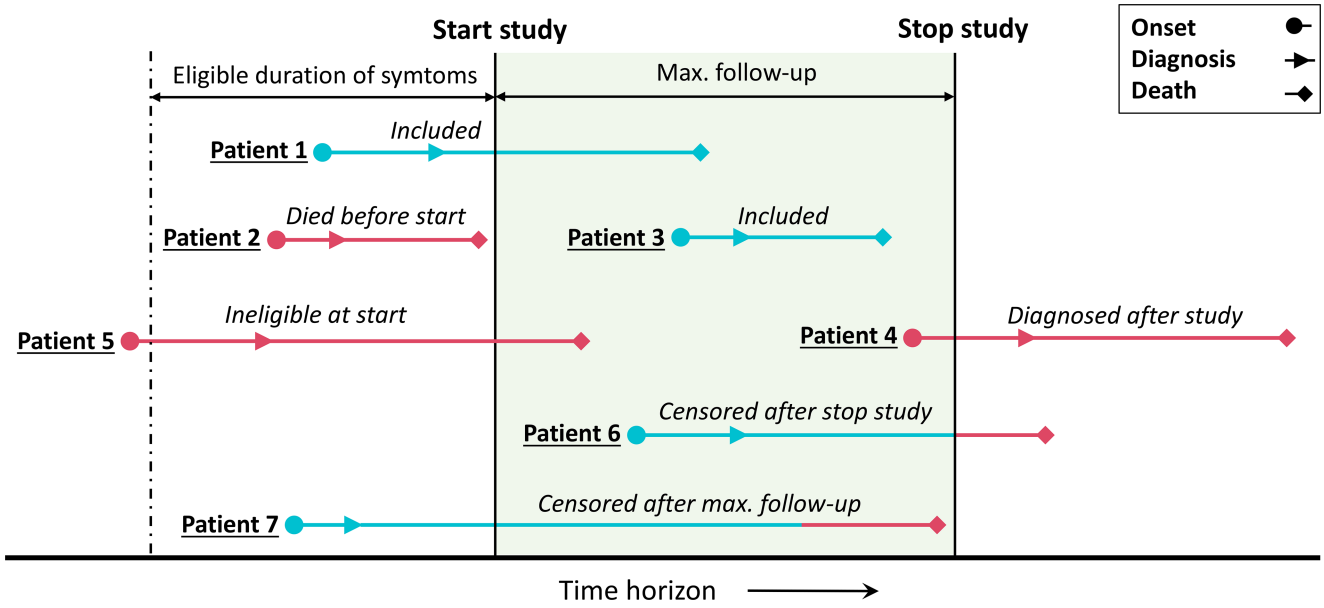


Figure 1 Identifying the eligible trial population and external nonparticipating patients. Illustration of different patient scenarios. The green area reflects the time window in which the trial was active. In red, the patients who were ineligible or whose follow-up data were partially ineligible for use as external control data. The blue patients define the eligible trial population and were used for propensity matching.

have been considered for study participation. As such, we applied an additional selection criterion to exclude those patients with a symptom duration of more than 36 months at the start date of the trial (November 2008). Although the symptom duration might have been less than 36 months at diagnosis, these patients would never have been enrolled in the trial, as they would have failed the 36-month eligibility criterion at the time of screening (illustrated as patient 5). Finally, we harmonized the censoring distribution between the registry and the trial by administratively censoring registry patients after the trial was completed (patient 6), or when a patient reached the maximum theoretical follow-up time in the trial (patient 7).

Matching external control patients to trial participants

After defining the eligible population, we matched eligible external control patients one-to-one with trial participants using propensity scores. The propensity score represents the conditional probability of belonging to a particular group, in this case, being a trial participant, given a set of baseline characteristics. Propensity scores were derived from a logistic regression model including duration of symptoms, age, body mass index (BMI), %predicted FVC, domains of the ALS functional rating scale (ALSFRS-R; i.e., bulbar, fine motor, gross motor, respiratory), rate of change in ALSFRS-R total score (ΔFRS: ALSFRS-R total score−48, divided by duration of symptoms),[23] El Escorial classification,[20] sex, site of symptom onset, and the difference between date of diagnosis and trial start. Continuous variables were modeled using restricted cubic splines to allow for potential nonlinear relationships. Nearest neighbor matching was used by computing the difference in propensity scores between each trial participant and each external control patient.

Statistical analysis

In total, 11% of the data, regarding the key inclusion criteria for determining eligibility, were missing from the registry at the time of diagnosis. To identify eligible patients in the registry, missing data were addressed by creating multiple imputed datasets ($n = 100$). The imputation model included all covariates and took into account additional information on neurological as well as laboratory examinations; survival time was modeled using the Nelson-Aalen estimator (cumulative hazard rate). Subsequently, in each imputed dataset, we identified all eligible patients, fitted the propensity model, and matched trial participants to eligible external control patients. Results across different imputations were pooled using Rubin's rules[24]; Kaplan–Meier curves across imputations were pooled based on a complementary log–log transformation.[25]

To determine whether the matched external control patients were suitable to augment the trial participants allocated to placebo (randomized controls), we defined an equivalence test to compare the restricted mean survival time during the trial period (RMST; i.e., the area under the cumulative survival curve).[26] Equivalence was deemed justified if the point estimate of the RMST of the matched external controls fell within the 80% confidence interval (CI) of the RMST of the randomized controls and, vice versa, the point estimate of the randomized controls within the 80% CI of the matched external controls.[14] If the equivalence criterion was met, a Cox proportional hazards model would be used to compare the patients randomized to lithium carbonate with the combined control population (randomized controls plus the matched external controls). If the equivalence criterion was not met, the comparison would be restricted to the randomized treatment and control arms. We conducted a sensitivity analysis applying the original event definition used in the clinical trial (death or respiratory insufficiency). Moreover, we performed an analysis with a nonconcurrent cohort in the registry by shifting the hypothetical runtime of the trial and repeating all selection and matching steps as described above. As such, this changed the calendar period from which external controls were borrowed, but kept other elements constant (e.g., maximum follow-up time and censoring mechanism in survival end point). Specifically, the nonconcurrent cohort was defined as all patients in the registry who had been diagnosed and who were alive between September 2003 and April 2006. In this period, patients were included in the dataset by referral (**Figure S2**),[21] where referral-based cohorts tend to have a better overall survival, as patients must survive at least until referral.[27]

The original clinical trial was based on a fully sequential design,[28] an interim analysis being conducted after every two to three events. To enable a comparison with the original design, and illustrate the methodology for the sequential analysis, we defined a group-sequential procedure with 20 planned interim analyses. To control type 1 and 2 errors, both superiority and futility boundaries were calculated based on Kim-DeMets alpha- and beta-spending functions (*rho* of 3; resembling conservative

O-Brien-Fleming type boundaries). The full design is presented in **Table S1** and was specified according to the original trial protocol. At each interim analysis–conducted after reaching a prespecified number of events in the RCT–we identified all eligible external patients who had been diagnosed before the interim analysis date and used their follow-up information up to the date of the interim analysis. Subsequently, we repeated the matching process, equivalence testing, and treatment effect estimation. Interim analyses were planned based on the accumulating information combining trial participants and matched external control patients, where the information fraction at a certain time was recalculated as (# of events in the treatment arm) × (# of events in the control arm)/(# of events in all patients) divided by the required information to detect an HR of 0.56 with a one-sided alpha of 5% and power of 90%.[29]

RESULTS

A total of 1,141 patients were identified in the registry who had been diagnosed and who were alive during the runtime of the trial. Of these, 473 (41.5%) fulfilled the inclusion criteria and would have been eligible for trial participation (**Figure 2**). Of all 473 eligible patients, 133 (28.1%) ultimately participated and 340 (71.9%) did not. Their patient characteristics are provided in **Table 1**. Compared to trial participants, eligible nonparticipating patients were older (6.2 years, 95% CI: 3.9–8.4), had a more aggressive phenotype, as reflected by their ΔFRS and duration of symptoms, and overall had a poorer health, as reflected by the directional differences in ALSFRS-R total score, %predicted FVC and BMI.

Using propensity matching, we identified 133 patients among the 340 eligible nonparticipating patients who were 1:1 matched with the trial participants and formed the matched external control group. After matching, there were no statistically significant differences in baseline characteristics between the trial participants and the matched external controls (**Table 1**). The distribution of the propensity scores is provided in **Figure S3**. **Table S2** provides the characteristics of the trial participants, stratified by the absolute difference in predicted probability with their match. Trial participants with a poorer match comprised young patients with a slowly progressive disease; they were under-represented among the eligible nonparticipating patients.

Differences in overall cumulative survival between trial participants allocated to placebo (randomized controls) and all 340 eligible nonparticipating patients are provided in **Figure 3a**. Overall, prior to propensity matching, the criterion for equivalence was not met, violating the assumption that eligible nonparticipating patients are equal in their RMST compared with randomized controls. After propensity matching, equivalence was established between the randomized and 133 matched external controls, resulting in a comparable cumulative survival (**Figure 3b**).

Impact on the treatment effect and trial design

In the original clinical trial, the HR between patients randomized to placebo and patients randomized to lithium carbonate was 1.09 (95% CI: 0.65–1.85), in favor of placebo. By supplementing the randomized controls with 133 propensity matched external controls in the expanded control arm, a near identical effect size was obtained together with a 17.4% reduction in the confidence interval width (**Figure 4**): the estimated HR was 1.11 (95% CI: 0.72–1.71). Although a larger reduction could be achieved by including all 340 eligible nonparticipating patients, the mismatch in overall survival with trial participants affects the treatment effect estimate and potentially introduces nontrivial bias (all eligible concurrent, **Figure 4**). Here, the bias shifts the effect estimate in favor of lithium carbonate as eligible nonparticipating patients had a poorer survival. A similar but reversed bias can be observed by including an eligible population that is nonconcurrent with the trial. In this case, the nonconcurrent external patients were identified by referral in the early days of the registry,[21] resulting in a selection toward long survivors. As this was a latent selection mechanism, it could not be addressed by propensity matching. Finally, we illustrate the impact of using the original event

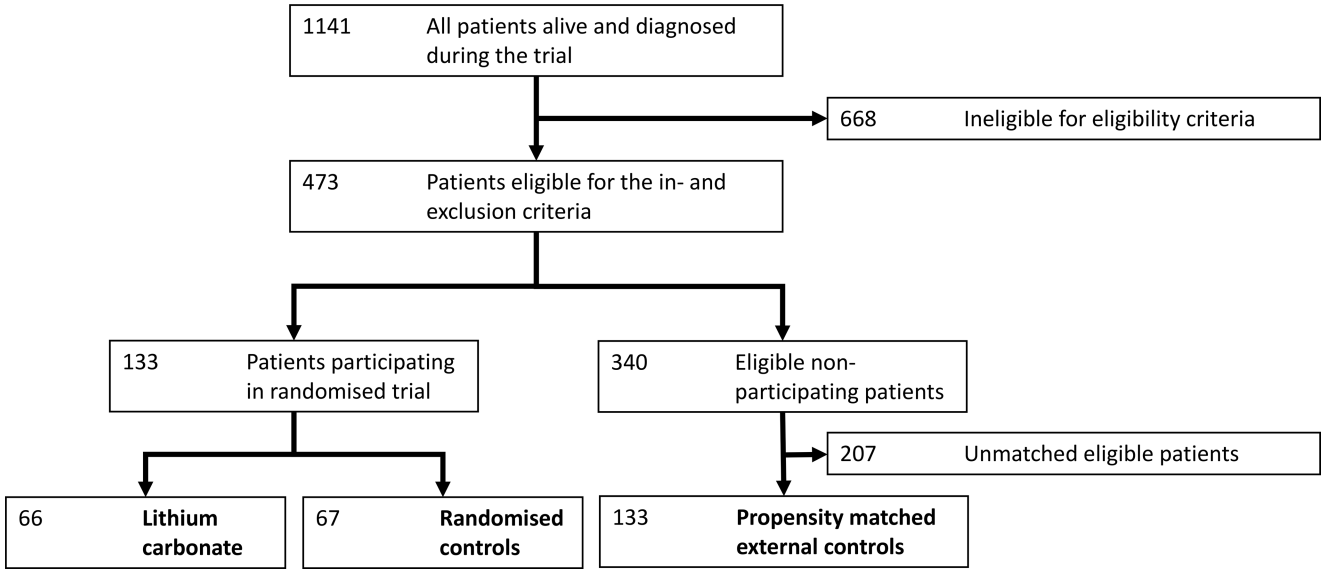


Figure 2 Flowchart of patient populations. All patients alive in the registry were identified, 340 of whom did not participate in the clinical trial, but were eligible for its in- and exclusion criteria. This population was used to identify 133 external controls that were propensity matched with all 133 trial participants.

Table 1 Patient characteristics of the general, eligible, participating and matched populations

Characteristic	All patients alive during the trial (*N* = 1,141)	Eligible for in- and exclusion criteria					
		All eligible patients (*N* = 473)	Patients participating in trial (*N* = 133)	Eligible nonparticipating patients		*P* value trial participants vs. all nonparticipants	*P* value trial participants vs. matched nonparticipants
				All nonparticipants (*N* = 340)	Matched nonparticipants (*N* = 133)		
Age, years	62 (12)	62 (12)	58 (12)	64 (11)	59 (11)	<0.001	0.47
Sex, male	677 (59%)	281 (59%)	80 (60%)	201 (59%)	77 (58%)	0.92	0.74
Onset, spinal	806 (71%)	333 (70%)	95 (71%)	238 (70%)	97 (73%)	0.84	0.78
El Escorial, clinically definite	221 (19%)	105 (22%)	34 (26%)	71 (21%)	34 (26%)	0.20	0.60
Symptom duration, months[a]	11.6 (11.4)	11.4 (7.0)	12.4 (10.0)	10.9 (6.5)	12.0 (7.9)	0.016	0.63
FVC, %predicted	89 (21)	94 (16)	95 (17)	93 (15)	96 (15)	0.175	0.90
ALSFRS-R total score	40 (6)	40 (5)	41 (5)	40 (5)	40 (4)	0.33	0.93
ΔFRS, points per month[a]	−0.58 (0.73)	−0.56 (0.57)	−0.50 (0.51)	−0.59 (0.60)	−0.50 (0.50)	0.044	0.62
Body mass index, kg/m^2	24.8 (4.0)	24.8 (4.0)	25.4 (4.5)	24.6 (3.8)	25.4 (4.1)	0.081	0.98

Data are expressed as mean (SD) or *n* (%).
ALSFRS-R, amyotrophic lateral sclerosis functional rating scale–revised; FVC, forced vital capacity; ΔFRS, ALSFRS-R–48/symptom duration.
[a]Median (interquartile range).

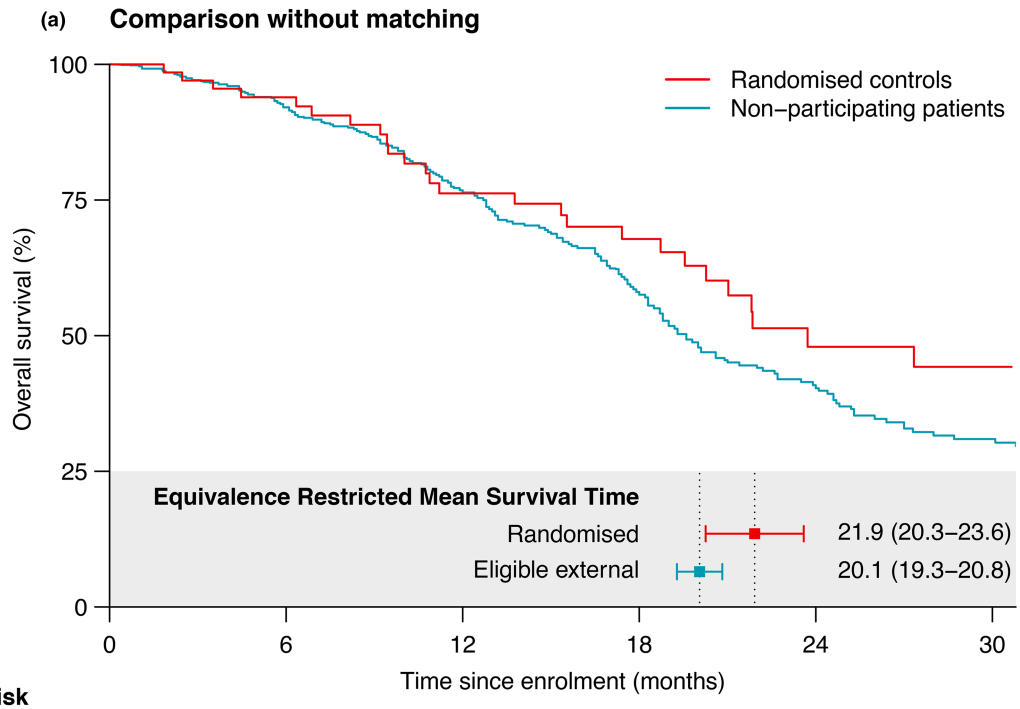
definition (composite event, **Figure 4**). For the original trial, this had only a minor impact on the outcome (**Figure S1**), resulting in nearly the same effect size for lithium. However, as the respiratory components were missing from the registry, adding the external data introduces some bias in favor of control. Significantly, the test for equivalence was not violated, illustrating its limitations for detecting small, yet potentially nontrivial outcome differences between randomized and external controls.

Figure 5 depicts the interim analysis scheme and the evolution of the HR over time during the trial, with and without addition of propensity matched external controls. The original trial crossed the futility border on May 24, 2011. By adding propensity matched external controls, the information required for each interim analysis was reached sooner, resulting in earlier and more frequent interim analyses. The augmented clinical trial crossed the futility border on October 7, 2010, 7.5 months earlier than the original design. The sample sizes at each interim are depicted in **Figure S4**. These findings indicate that augmenting the randomized controls with propensity matched external controls could avoid unnecessary randomization of 17 patients (12.8%) and reduce the overall trial duration by 25.0% for a futile trial.

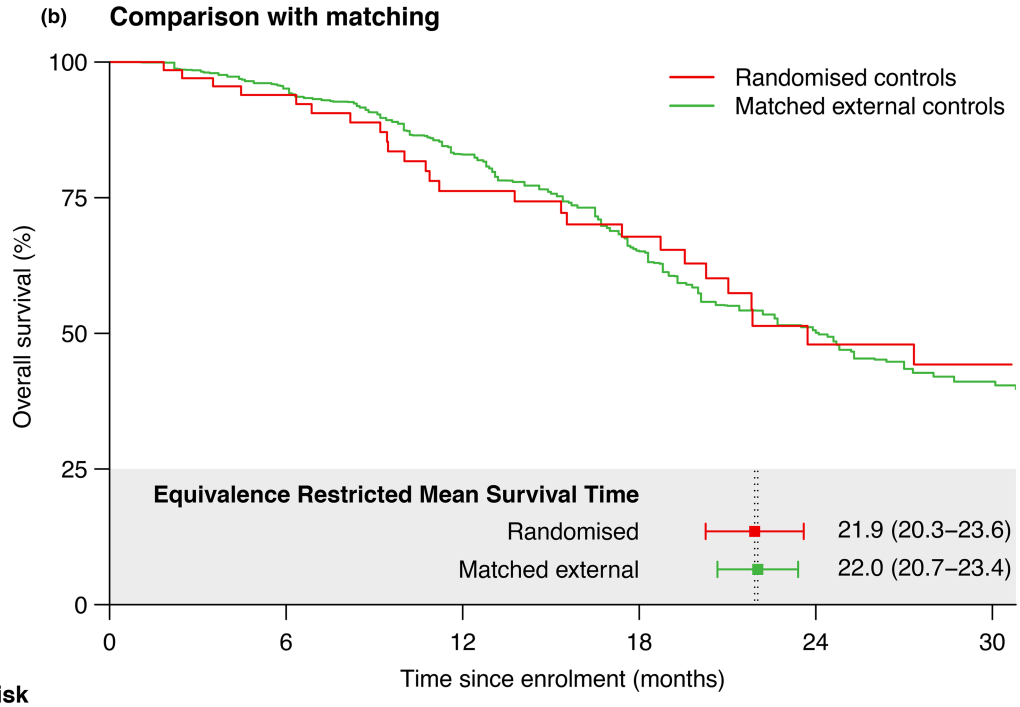
DISCUSSION

In this study, we leveraged a high-quality, patient-level concurrent registry to improve a clinical trial in ALS by augmenting the comparator group. We successfully identified a cohort of patients in the registry to serve as concurrent controls, who were interchangeable with patients participating in the clinical trial. By integrating these external controls into a hybrid design, we improved precision of the clinical trial, reduced the time to reach a decision, and lowered the number of randomized patients and those allocated to placebo. Population-based registries, such as the one we used, are strong candidates as external sources of controls for hybrid clinical trials; they can limit significant sources of bias, such as differing standards of care and mismatches in calendar time. Especially for diseases with a rare and significant unmet medical need, with limited treatment options available, the hybrid design allows studies to be conducted when large placebo arms are unethical or infeasible.

The primary obstacle of integrating external data into clinical trials is the potential introduction of bias that may lead to either a false-negative or false-positive conclusion.[30] Geographic and temporal differences between the trial population and the external dataset could have a significant impact on trial end points.[13] Some of these factors may be easily identifiable, such as difference in standard of care, but could also be more subtle, such as differences in referral patterns, health delivery, or cultural norms regarding experimental treatments,[8,12] or differences in genetic, biological, or pathophysiological disease mechanisms.[31] These differences may be further aggravated over time and are difficult to capture, resulting in residual bias despite the use of rigorous statistical methodology. We addressed these challenges by borrowing information from a parallel, perpetual, population-based registry that enrolls patients within the same geographic area and calendar period as the clinical trial. Nevertheless, other sources of bias may remain, including a placebo effect in the randomized control arm that is absent in the



Number at risk (number censored)						
Randomised	67 (0)	56 (7)	41 (12)	29 (20)	14 (28)	2 (39)
Non–participants	340 (0)	277 (37)	186 (85)	117 (112)	70 (126)	47 (134)



Number at risk (number censored)						
Randomised	67 (0)	56 (7)	41 (12)	29 (20)	14 (28)	2 (39)
Matched external	133 (0)	111 (15)	81 (32)	53 (44)	36 (50)	25 (55)

Figure 3 Overall survival of eligible nonparticipating patients, randomized controls and matched external controls. Equivalence between control populations in their overall survival, defined as the time from enrollment to death from any cause, was based on the 80% confidence interval around the restricted mean survival time (i.e., the area under the cumulative survival curve).

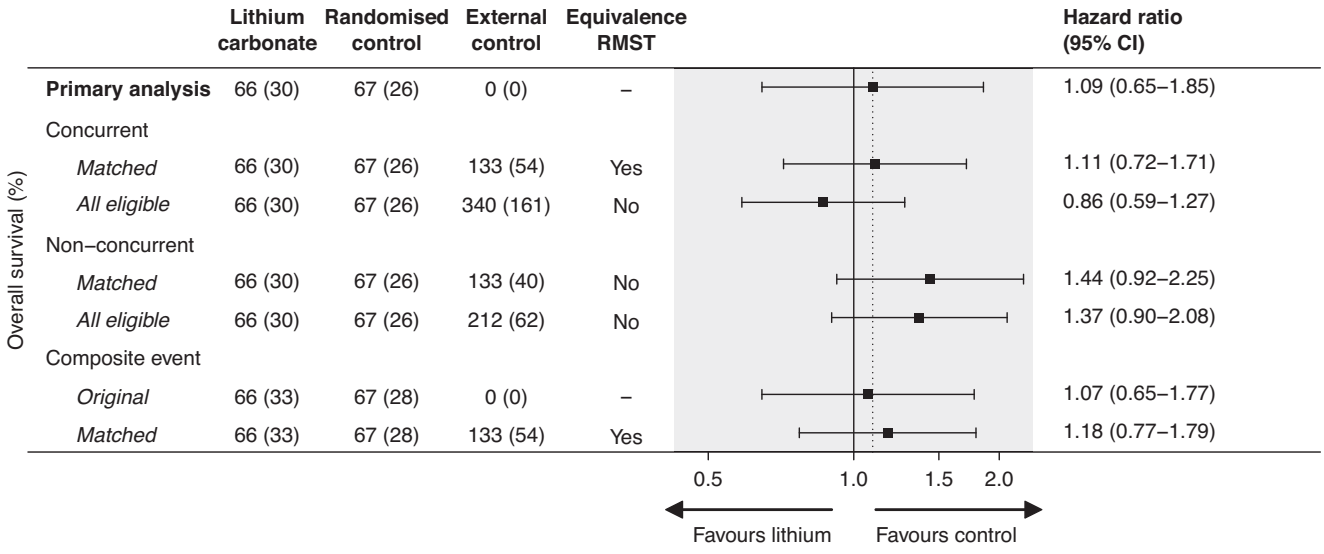


Figure 4 Impact of external controls on the effect of lithium carbonate. CI, confidence interval; RMST, restricted mean survival time. Numbers for each group reflect the number of patients (number of events).

external controls.[8] This may be of lesser significance for "definite" end points like mortality, but could play a nontrivial role for more subjective outcomes, such as physical functioning or quality of life. The equivalence comparison between outcomes of randomized and external controls helps to identify these potential differences, but dedicated studies are needed to assess how such residual bias may affect the operating characteristics of hybrid designs.

The proposed methodology is limited to clinical trials that are enrolling patients in geographic areas where such registries exist. Fortunately, the number of population-based registries is expanding rapidly and their initiation has been actively encouraged by regulatory agencies.[2] For ALS, 22 registries were recently identified; many of these are operating in countries with active clinical trials.[32] Nevertheless, attention should be given to the gaps in registry practice, which may not align with the way trials are evolving. More than just basic diagnostic details are needed for better utilization of RWD.[8] By linking disease registries with national personal records and administrative databases, collection of primary outcomes, such as date of death, can be collected in near real-time. A key challenge is the collection of harmonized, standardized, longitudinal disease-specific outcomes, such as vital capacity, hospitalization, or quality of life on an unbiased and population-based scale. Recent advancements in digital healthcare technology could be of major help to address these obstacles,[16] but scalability to a population-based level would be required for future global hybrid clinical trials.

Besides the quality of the registry, it is equally important for a sponsor to prespecify their intention to use RWD. This requires appropriate selection of end points that are readily available in the RWD, preparation of a detailed data collection, and analytical plan at the outset of the study.[33] Sponsors should make explicit the intended data source, the selection and matching methods, how external data will be borrowed during the statistical analysis, and define contingency and exit strategies when matching is unsuccessful. This may require feasibility assessments of the registry, together with refinement of the study's eligibility criteria and the primary end point, taking into account the information available in the registry. Moreover, although hybrid approaches have already been used successfully in regulatory decisions,[2,4] it is important to continue engagement with the major regulatory bodies, in order to assure conducting a study according to the regulatory standards.

Our study has several limitations. First, the original trial primary end point was defined as time to death or respiratory insufficiency, requiring prospective registration of respiratory events. In our registry, such level of detail was missing and required a change in primary outcome to death alone. In our case, it had little impact on the outcome (**Figure S1**), but could potentially bias study findings (**Figure 4**); this underscores the importance of prospectively considering end point definitions in hybrid clinical trials. Second, our results reveal non-negligible differences between eligible patients in the registry and those who participate in clinical trials.[17] We used propensity scores to address these differences and show how a matched cohort with equivalent survival could be identified in the registry. We also illustrated that, for patients who are over-represented in clinical trials, it is more difficult to find a suitable match in the registry (**Table S1**), and unmeasured confounders may not have been adequately addressed. To counteract these limitations, we implemented a gatekeeping equivalence test using the randomized controls. This is a major strength of the hybrid design and a critical mechanism to prevent the wrong conclusion being drawn; a mechanism that is missing from single-armed trials. Detailed studies are still required to provide guidance on how to handle poorly matched individuals.

For prospective use of the hybrid design, one can either design the trial considering the additional external controls, thereby actively planning a smaller randomized control group, or design the trial with a conventional randomized arm and increase statistical power by adding the additional external controls. As an illustration, in order to detect a HR of 0.56 in a conventional, 1:1 randomized trial with 80% power, 298 patients are required, assuming our observed event rates.[34] In a hybrid design, the treated

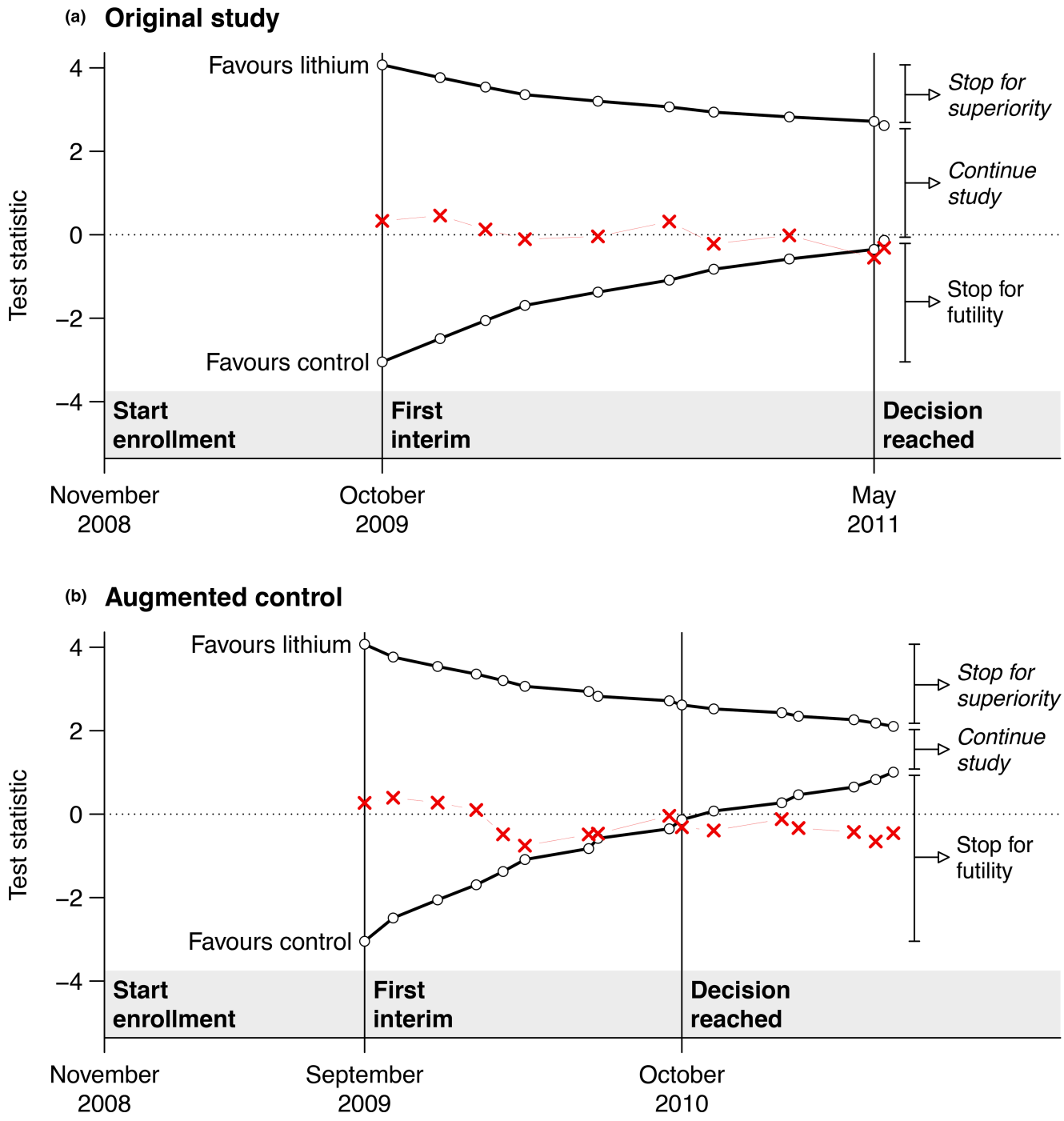


Figure 5 Development of the treatment effect over time. The black solid lines reflect the superiority and futility boundaries that define the decision criteria used at each interim analysis. As soon as the standardized test statistic for the treatment effect (red crosses) falls outside one of the boundaries, the study can be stopped for (in)effectiveness. (**a**) reflects the original study design, utilizing only randomized controls, whereas (**b**) reflects the augmented design with propensity matched external controls. The numerical values are provided in **Table S1**.

to control ratio is unbalanced, namely: one treated vs. one randomized control plus 2 external controls. Given this unbalanced design, instead of 298 patients, 356 patients are required to maintain 80% power.[29] As 50% of the patients are "borrowed" from the registry, the planned sample size is 178 (a 40.2% reduction compared with the conventional design). The risk is that when one fails to find suitable external controls, power will be reduced to 58%. Alternatively, one can plan for a conventional design and add the external controls. Here, adding 298 external controls to the conventional design increases power from 80% to 95% or – depending on the event distribution – allows for a reduction in trial duration.[35] Of note, the above example does not take into account the required inflation of the sample size – or reduction in statistical power – when introducing interim

analyses into the design. This applies for designs with and without external controls. In general, one or two conservative interim analyses usually have a minor impact on sample size, but this may change when using more liberal stopping rules or increasing the number of interim analyses.

Other considerations include the timing for identifying matched external controls, which should be carefully planned prospectively. A reasonable number of trial participants must be enrolled to generate reliable propensity scores between the trial and registry. An area of investigation is the prospective matching during interim analyses (e.g., are all patients rematched at every interim analysis or only the newly enrolled patients), where the latter may be preferred. In addition, the number of interim analyses should be limited, not only as adding more interim analyses may only bring modest efficiency gains, but each interim analysis will have a significant operational impact in order to prepare and update the registry information. Moreover, one could consider the use of different statistical strategies.[8,30,36–38] In this study, we matched trial participants individually with external controls; the major benefit is that this allows for full transparency of the matching process. Combined with the equivalence test of outcomes, one further ensures exchangeability of the randomized and external controls. A downside is that parts of the external data are disregarded, matched external controls receive equal weight compared with randomized controls, and one assumes that baseline characteristics explain all the differences. Other strategies include the construction of prior distributions, such as weighted log-likelihood methods using a power prior, and hierarchical models based on meta-analytical predictive priors. These methods estimate the heterogeneity between randomized and external controls and account for it by discounting the degree of borrowing; but they do require assumptions to be made about the prior distributions.[8,30,38] Further comparison of these strategies is warranted.

In conclusion, in this study, we have shown the feasibility of conducting a hybrid clinical trial in ALS with both a randomized control arm and propensity matched external controls. The design preserves key features of RCTs, such as randomization, while benefiting from RWD by augmenting the randomized control group. A concurrent patient-level registry with rigorous matching may be a strong candidate to serve as the source for the external controls. Our approach minimizes bias due to a mismatch in calendar time and differences in standard of care, and may ultimately accelerate the development of new treatments.

SUPPORTING INFORMATION

Supplementary information accompanies this paper on the *Clinical Pharmacology & Therapeutics* website (www.cpt-journal.com).

ACKNOWLEDGMENTS

The authors thank Brenda Vollers-King for assistance with language editing.

FUNDING

This work was funded by the UCB-Stanford Digital Health Research Collaborative. The Sponsor of this study was not involved in the study design; in the collection and analysis; nor in the decision to submit the paper for publication. Scientific employees of the Sponsor contributed to the interpretation of study findings and writing the report.

CONFLICTS OF INTEREST

The authors declared no competing interests for this work.

AUTHOR CONTRIBUTIONS

R.P.A.v.E., L.H.v.d.B., K.C.B.R., L.T., T.L.L., L.M.N., C.L., A.S., J.G.-S., and Y.L. wrote the manuscript. R.P.A.v.E., L.H.v.d.B., K.C.B.R., L.T., T.L.L., L.M.N., C.L., A.S., J.G.-S., and Y.L. designed the research. R.P.A.v.E. and Y.L. performed the research. R.P.A.v.E. and Y.L. analyzed the data. R.P.A.v.E., C.L., and Y.L. contributed new reagents/analytical tools.

DATA AVAILABILITY STATEMENT

De-identified and simulated participant data can be requested by any qualified researcher by contacting the corresponding authors. The statistical codes underlying this publication were made open-source available as R-package ("ALSRWDHybridDesign"; https://github.com/CISD-Stanford/ALSRWDHybridDesign).

1. Avorn, J. The $2.6 billion pill–methodologic and policy considerations. *N. Engl. J. Med.* **372**, 1877–1879 (2015).
2. Jansen-van der Weide, M.C. *et al*. Rare disease registries: potential applications towards impact on development of new drug treatments. *Orphanet J. Rare Dis.* **13**, 154 (2018).
3. Arlett, P., Kjaer, J., Broich, K. & Cooke, E. Real-world evidence in EU medicines regulation: enabling use and establishing value. *Clin. Pharmacol. Ther.* **111**, 21–23 (2022).
4. Purpura, C.A., Garry, E.M., Honig, N., Case, A. & Rassen, J.A. The role of real-world evidence in FDA-approved new drug and biologics license applications. *Clin. Pharmacol. Ther.* **111**, 135–144 (2022).
5. Food Drug Administration Center for Drugs Evaluation Research Complex innovative trial design meeting program. <https://www.fda.gov/drugs/development-resources/complex-innovative-trial-design-meeting-program> (2022). Accessed December 27, 2022.
6. Food Drug Administration Center for Drugs Evaluation Research Complex innovative trial designs. <https://fda.report/media/129256/Complex+Innovative+Trial+Designs+Brochure_508.pdf> (2022). Accessed December 27, 2022.
7. European Medicines Agency Accelerating Clinical Trials in the EU: Workplan 2022–2026. <https://www.ema.europa.eu/en/documents/other/act-eu-multi-annual-workplan-2022-2026_en.pdf> (2022). Accessed December 27, 2022.
8. Hall, K.T. *et al*. Historical controls in randomized clinical trials: opportunities and challenges. *Clin. Pharmacol. Ther.* **109**, 343–351 (2021).
9. Ghadessi, M. *et al*. A roadmap to using historical controls in clinical trials–by drug information association adaptive design scientific working group (DIA-ADSWG). *Orphanet J. Rare Dis.* **15**, 69 (2020).
10. Zhou, Y., Wang, F., Tang, J., Nussinov, R. & Cheng, F. Artificial intelligence in COVID-19 drug repurposing. *Lancet Digit Health* **2**, e667–e676 (2020).
11. Jeon, J.Y. *et al*. Development of an external control arm using electronic health record-based real-world data to evaluate the efficacy of COVID-19 treatment. *Clin. Pharmacol. Ther.* **113**, 1274–1283 (2023).
12. Burger, H.U. *et al*. The use of external controls: to what extent can it currently be recommended? *Pharm. Stat.* **20**, 1002–1016 (2021).
13. Rahman, R. *et al*. Leveraging external data in the design and analysis of clinical trials in neuro-oncology. *Lancet Oncol.* **22**, e456–e465 (2021).

14. Yuan, J., Liu, J., Zhu, R., Lu, Y. & Palm, U. Design of randomized controlled confirmatory trials using historical control data to augment sample size for concurrent controls. *J. Biopharm. Stat.* **29**, 558–573 (2019).
15. Hernan, M.A., Sauer, B.C., Hernandez-Diaz, S., Platt, R. & Shrier, I. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. *J. Clin. Epidemiol.* **79**, 70–75 (2016).
16. Stern, A.D. *et al*. Advancing digital health applications: priorities for innovation in real-world evidence generation. *Lancet Digit Health* **4**, e200–e206 (2022).
17. Chio, A. *et al*. ALS clinical trials: do enrolled patients accurately represent the ALS population? *Neurology* **77**, 1432–1437 (2011).
18. Verstraete, E. *et al*. Lithium lacks effect on survival in amyotrophic lateral sclerosis: a phase IIb randomised sequential trial. *J. Neurol. Neurosurg. Psychiatry* **83**, 557–564 (2012).
19. Fornai, F. *et al*. Lithium delays progression of amyotrophic lateral sclerosis. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 2052–2057 (2008).
20. Brooks, B.R., Miller, R.G., Swash, M., Munsat, T.L. & World Federation of Neurology Research Group on Motor Neuron Diseases El Escorial revisited: revised criteria for the diagnosis of amyotrophic lateral sclerosis. *Amyotroph. Lateral Scler. Other Motor Neuron Disord.* **1**, 293–299 (2000).
21. de Jongh, A.D. *et al*. Incidence, prevalence and geographical clustering of motor neuron disease in The Netherlands. *Neurology* **96**, e1227–e1236 (2021).
22. Pocock, S.J. The combination of randomized and historical controls in clinical trials. *J. Chronic Dis.* **29**, 175–188 (1976).
23. Kimura, F. *et al*. Progression rate of ALSFRS-R at time of diagnosis predicts survival time in ALS. *Neurology* **66**, 265–267 (2006).
24. Rubin, D.B. & Schenker, N. Multiple imputation in health-care databases: an overview and some applications. *Stat. Med.* **10**, 585–598 (1991).
25. Morisot, A., Bessaoud, F., Landais, P., Rébillard, X., Trétarre, B. & Daurès, J.P. Prostate cancer: net survival and cause-specific survival rates after multiple imputation. *BMC Med. Res. Methodol.* **15**, 54 (2015).
26. Lu, Y. & Tian, L. Statistical considerations for sequential analysis of the restricted mean survival time for randomized clinical trials. *Stat. Biopharm. Res.* **13**, 210–218 (2021).
27. Rooney, J.P.K., Brayne, C., Tobin, K., Logroscino, G., Glymour, M.M. & Hardiman, O. Benefits, pitfalls, and future design of population-based registers in neurodegenerative disease. *Neurology* **88**, 2321–2329 (2017).
28. Whitehead, J. *The Design and Analysis of Sequential Clinical Trials* Revised 2nd edn. (J. Wiley & Sons, Chichester; New York, 1997).
29. Wu, J. *Statistical Methods for Survival Trial Design: With Applications to Cancer Clinical Trials Using R* (Taylor & Francis, Boca Raton, FL, 2018).
30. van Rosmalen, J., Dejardin, D., van Norden, Y., Löwenberg, B. & Lesaffre, E. Including historical data in the analysis of clinical trials: is it worth the effort? *Stat. Methods Med. Res.* **27**, 3167–3182 (2018).
31. Wang, R.S., Croteau-Chonka, D.C., Silverman, E.K., Loscalzo, J., Weiss, S.T. & Hall, K.T. Pharmacogenomics and placebo response in a randomized clinical trial in asthma. *Clin. Pharmacol. Ther.* **106**, 1261–1267 (2019).
32. Cook, S.F. *et al*. A descriptive review of global real world evidence efforts to advance drug discovery and clinical development in amyotrophic lateral sclerosis. *Front. Neurol.* **12**, 770001 (2021).
33. Baumfeld Andre, E., Reynolds, R., Caubel, P., Azoulay, L. & Dreyer, N.A. Trial designs using real-world data: the changing landscape of the regulatory approval process. *Pharmacoepidemiol. Drug Saf.* **29**, 1201–1212 (2020).
34. Schoenfeld, D.A. & Richter, J.R. Nomograms for calculating the number of patients needed for a clinical trial with survival as an endpoint. *Biometrics* **38**, 163–170 (1982).
35. van Eijk, R.P.A. *et al*. Critical design considerations for time-to-event endpoints in amyotrophic lateral sclerosis clinical trials. *J. Neurol. Neurosurg. Psychiatry* **90**, 1331–1337 (2019).
36. Kopp-Schneider, A., Calderazzo, S. & Wiesenfarth, M. Power gains by using external information in clinical trials are typically not possible when requiring strict type I error control. *Biom. J.* **62**, 361–374 (2020).
37. Liu, Y. *et al*. Matching design for augmenting the control arm of a randomized controlled trial using real-world data. *J. Biopharm. Stat.* **32**, 124–140 (2022).
38. Schmidli, H., Haring, D.A., Thomas, M., Cassidy, A., Weber, S. & Bretz, F. Beyond randomized clinical trials: use of external controls. *Clin. Pharmacol. Ther.* **107**, 806–816 (2020).