Check for updates

CPT: Pharmacometrics & Systems Pharmacology

ARTICLE OPEN ACCESS

A Pharmacometrics-Informed Trial Simulation Framework for Optimizing Study Designs for Disease-Modifying Treatments in Rare Neurological Disorders

Yevgen Ryeznik[1] | Ralf-Dieter Hilgers[2] | Nicole Heussen[2,3] | Emmanuelle Comets[4,5] | France Mentré[4] | Niels Hendrickx[4] | Mats O. Karlsson[1] | Andrew C. Hooker[1] | Alzahra Hamdan[1] | Xiaomei Chen[1] | Rebecca Schüle[6,7,8] | Matthis Synofzik[6,7] | Oleksandr Sverdlov[9]

[1]Department of Pharmacy, Uppsala University, Uppsala, Sweden | [2]Department of Medical Statistics, RWTH Aachen University, Aachen, Germany | [3]Medical School, Sigmund Freud Private University, Vienna, Austria | [4]Université Paris Cité and Université Sorbonne Paris Nord, Inserm, IAME, Paris, France | [5]Univ Rennes, Inserm, EHESP, Irset (Institut de Recherche en santé, Environnement et Travail) UMR_S 1085, Rennes, France | [6]Division Translational Genomics of Neurodegenerative Diseases, Hertie Institute for Clinical Brain Research (HIH), University of Tübingen, Tübingen, Germany | [7]German Center for Neurodegenerative Diseases (DZNE), Tübingen, Germany | [8]Division of Neurodegenerative Diseases, Department of Neurology, Heidelberg University Hospital, Germany | [9]Novartis Pharmaceuticals Corporation, East Hanover, New Jersey, USA

Correspondence: Oleksandr Sverdlov (alex.sverdlov@novartis.com)

Received: 28 February 2025 | **Revised:** 9 July 2025 | **Accepted:** 11 July 2025

Funding: This work was supported by the European Union, as part of the EJP RD WP20 Innovation Statistics consortium "EVIDENCE-RND" (to M.S., R.S., M.O.K, R.-D.H., and F.M.) from the European Union's Horizon 2020 research and innovation programme under the EJP RD COFUND-EJP (No. 825575); and later as part of the project European Rare Disease Research Alliance (ERDERA), GA no. 101156595, funded under call HORIZON-HLTH-2023-DISEASE-07 (to M.S., R.S., M.O.K., R.-D.H, and F.M.). R.-D.H. is coordinator and N.H. and O.S. are contributors of Realized supported by the Innovative Health Initiative Joint Undertaking (IHI JU) under grant agreement No. 101165912. The JU receives support from the European Union's Horizon Europe research and innovation programme and COCIR, EFPIA, Europa Bío, MedTech Europe, and Vaccines Europe.

Keywords: clinical design optimization | model-informed drug development | nonlinear mixed effects model | pharmacometrics | simulation

ABSTRACT

The development of new treatments for rare neurological diseases (RNDs) may be very challenging due to limited natural history data, lack of relevant biomarkers and clinical endpoints, small and heterogeneous patient populations, and other complexities. A systematic approach is needed for comparing various design and analysis strategies to identify "optimal" approaches for a clinical trial in a chosen RND with the given resource constraints. For this purpose, we propose a pharmacometrics-informed clinical scenario evaluation framework (CSE-PMx), which includes some important research hallmarks relevant to RND clinical trials: a disease progression model for simulating individual longitudinal outcomes, the choice of a suitable randomization method for trial design, and an option to perform subsequent statistical analysis with randomization tests. We illustrate the utility of CSE-PMx for an exemplary randomized trial to compare the disease-modifying effect of an experimental treatment versus control in patients with Autosomal-Recessive Spastic Ataxia Charlevoix Saguenay (ARSACS). In the considered example, our simulation evidence suggests that a nonlinear mixed-effects model (NLMEM) with a population-based likelihood ratio test analysis is valid, robust, and more powerful than some conventional methods such as two-sample *t*-test, analysis of covariance (ANCOVA), or a mixed model with repeated measurements (MMRM). Our proposed framework is very flexible and generalizable to clinical research in other rare disease indications.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2025 The Author(s). *CPT: Pharmacometrics & Systems Pharmacology* published by Wiley Periodicals LLC on behalf of American Society for Clinical Pharmacology and Therapeutics.

Summary

- What is the current knowledge on the topic?
  - The design of clinical trials of disease-modifying therapies for rare neurological diseases (RNDs) is challenging due to limited patient populations, sparsity of historical data, lack of reliable clinical endpoints, and other complexities. A systematic approach for designing and optimizing clinical trials of RNDs is warranted.
- What question does this study address?
  - A pharmacometrics-informed clinical scenario evaluation framework (CSE-PMx) is proposed for in silico evaluation and comparison of various statistical design and analysis strategies to assess the disease-modifying effect of an investigational therapy for a selected RND. Its utility is illustrated for a hypothetical clinical trial in patients with Autosomal-Recessive Spastic Ataxia Charlevoix Saguenay (ARSACS).
- What does this study add to our knowledge?
  - In silico evaluation of various clinical development options can help study investigators calibrate their study design for best performance. For an exemplary 1:1 Randomized Clinical Trial (RCT) of a disease-modifying therapy for ARSACS, a longitudinal design and analysis with a nonlinear mixed effects model (NLMEM) provides a valid and efficient procedure that has superior performance over some more conventional approaches.
- How might this change drug discovery, development, and/or therapeutics?
  - The proposed framework creates a new process for developing fit-for-purpose design and analysis strategies for clinical trials of disease-modifying therapies for RNDs. The approach may be generalizable to clinical research in other rare disease indications.

1 | Introduction

Over 300 million people worldwide are affected by rare diseases, and almost half of these diseases show neurological symptoms [1]. Rare neurological diseases (RNDs) present significant unmet medical needs. Many RNDs have a genetic basis, are likely to start in childhood, and last into adulthood; therefore, early detection and intervention are essential. For many RNDs, there is a lack of natural history data, relevant biomarkers, and clinical endpoints measuring treatment benefits and risks. Since RNDs involve small and heterogeneous patient populations, there is a great concern about how to design and conduct clinical trials for obtaining substantial evidence on safety and effectiveness for the approval of a new drug. Importantly, RND clinical trials must meet the same standards as those for more prevalent diseases [2].

To demonstrate robust evidence of disease-modifying treatment effect in a clinical trial, that is, the impact of a treatment to delay, slow, or reverse the progression of a disease by targeting its underlying cause, one should have a disease progression model that can reliably predict the individual disease trajectories with respect to the clinically meaningful outcome(s) and have proper statistical tools to quantify treatment effect and its significance. The traditional drug development paradigm, often requiring two large, randomized trials, is frequently not feasible in an RND setting where recruiting for even one such study is a challenge.

There is an increasing body of knowledge on methodologies for the design and analysis of RND clinical trials in small populations [3–6]. Pharmacometrics (PMx) methods that integrate drug-disease models are widely used in model-informed drug discovery and development [7, 8]. Combining various methodologies across quantitative disciplines can be synergistic and improve the probability of success of clinical development programs [9, 10]. Modeling and simulation methodologies are increasingly applied to calibrate clinical study design and analysis methods [11]. Systematic frameworks have been proposed for in silico evaluation and comparison of competing options for clinical programs, trial designs, and analysis methods to enable efficient decision-making and facilitate discussion with key stakeholders in the drug development enterprise [12, 13]. However, these frameworks may be insufficient for addressing some specific challenges of drug development for RNDs—the rarity of the diseases under consideration, the unique features of disease-modifying therapies, and the need to utilize drug–disease PMx models in the design and analysis of clinical trials. The current paper aims to propose a pharmacometrics-informed clinical scenario evaluation framework (CSE-PMx) for in silico evaluation and comparison of various clinical trial strategies to assess the disease-modifying effect of an investigational therapy for a selected RND. The Section 2 provides a general description of the CSE-PMx framework. The Section 3 showcases its application for a hypothetical 1:1 RCT in ARSACS patients. The Section 4 summarizes the added value of this work and outlines some possible extensions.

2 | Methods

A strategic question to be addressed with our CSE-PMx framework can be formulated as follows: which design options (combinations of randomization method, primary endpoint, and data analysis method) provide "optimal" strategies—valid, efficient, and robust methods for a clinical research study with the given resource constraints—to evaluate disease-modifying treatment effect in a selected patient population with a rare neurological disease? Our proposal extends the existing CSE framework [12], a structured approach used in the planning and assessment of clinical trials. To draw a parallel with the CSE framework, the above strategic question would fall in the category of "Trial Level CSE". However, our proposal includes several additional research hallmarks relevant to the RND clinical trials: a disease progression model based on available natural history data for simulating individual longitudinal outcomes, the choice of a suitable randomization method for trial design (assuming the study is randomized), and an option to perform subsequent statistical analysis with randomization tests. The latter option is important in view of that many RND disease-modifying therapies can be feasibly tested only in one adequate and well

21638306, 2025, 10, Downloaded from https://ascpt.onlinelibrary.wiley.com/doi/10.1002/psp4.70082 by Uppsala University Library, Wiley Online Library on [30/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

controlled trial, and it is essential that the proposed statistical method controls the probability of false-positive findings to meet the pre-requisites for regulatory approval.

Figure 1 shows the schematic of our proposed CSE-PMx framework. The first component, "Clinical Scenarios", includes a combination of Assumptions and Options that must be specified with the clinical team at the study planning stage. The second component, "Evaluation," includes a statistical simulation engine for simulating the design, conduct, and analysis of a clinical trial under a range of specified assumptions and options. The third component, "Performance Metrics," contains quantitative summaries from simulations. The last component, "Decision Making," includes the presentation of performance metrics of the considered options across various experimental scenarios in the format that would help the stakeholders make informed decisions on the choice of the fit-for-purpose design and analysis strategy. In what follows, we describe these components in more detail.

2.1 | Clinical Scenarios: Assumptions

2.1.1 | Disease-Specific Features

We distinguish two categories of disease-specific features: known and unknown. The "known" features are available to the investigators based on the existing data, for example, the disease progression model. The "unknown" features are the parameters for which no reliable information exists and whose values must be hypothesized, for example, the treatment effect of an experimental therapy. For many RNDs, there is a strong need to develop disease-modifying treatments. To obtain evidence that an investigational treatment slows down or halts disease progression, one should have a reliable disease progression model in the untreated population of patients. We assume that such a model is available to the investigator. As regards the treatment effect, it must be hypothesized based on clinical expert opinion. Ideally, a range of treatment effects should be explored, from no effect to an upside case that the treatment fully inhibits the progression of disease.

2.1.2 | Clinical Trial Parameters

Key clinical trial parameters in the framework include the feasible sample size, trial duration, and frequency of data collection.

2.1.3 | Additional Complexities

The framework can also account for additional complexities such as patient drop-out rates and the potential for experimental biases such as chronological bias, selection bias, etc. [14].

2.2 | Clinical Scenarios: Options

By "options" we mean different strategies for the design and analysis of a clinical trial (i.e., combinations of randomization method, primary endpoint, data analysis methodology, etc.) to reliably address study objectives that may be formulated as testing some clinical research hypothesis of disease-modifying treatment effect.

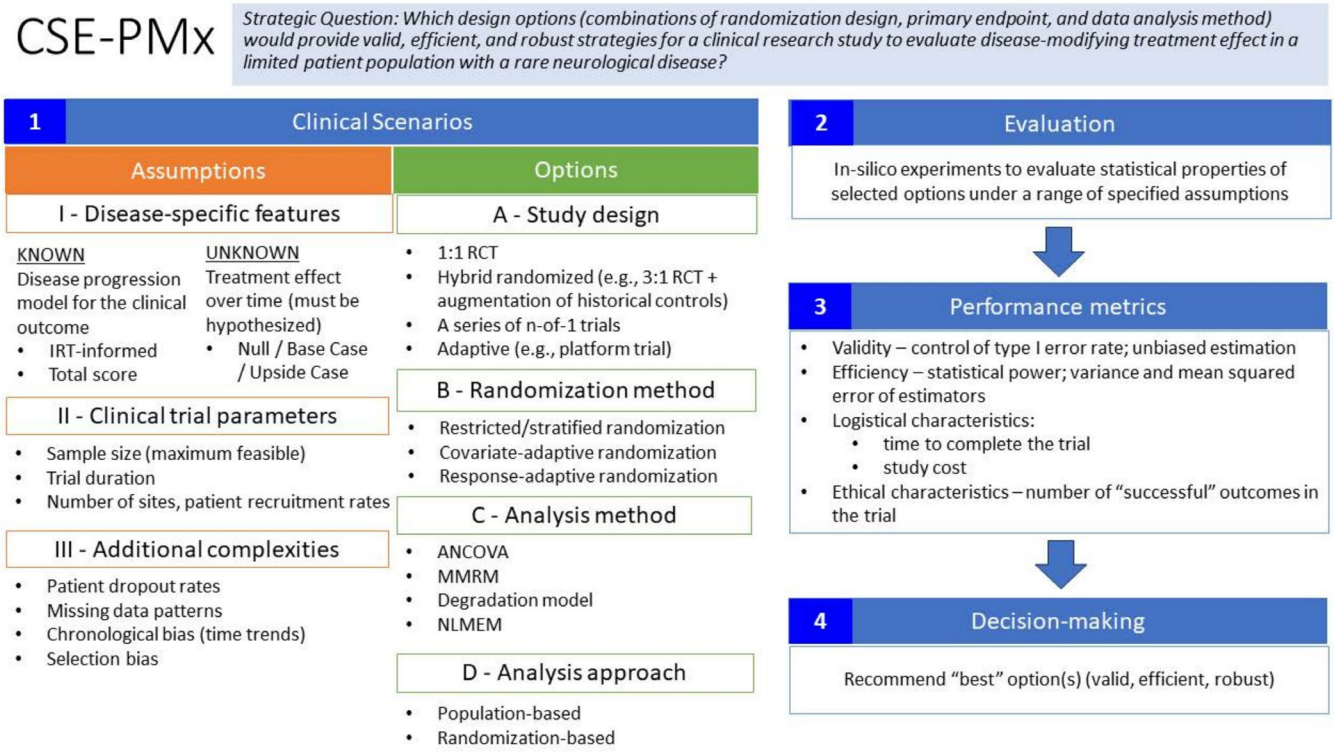

FIGURE 1 | Pharmacometrics-Informed Clinical Scenario Evaluation framework (CSE-PMx). ANCOVA = analysis of covariance, IRT = item response theory, MMRM = mixed model with repeated measurements, NLMEM = nonlinear mixed effects model, RCT = randomized controlled trial.

21638306, 2025, 10, Downloaded from https://ascpt.onlinelibrary.wiley.com/doi/10.1002/psp4.70082 by Uppsala University Library, Wiley Online Library on [30/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2.2.1 | Study Design

The choice of a study design is determined by the primary study objective. The "gold standard" research design is 1:1 randomized controlled trial (RCT). However, in an RND setting, 1:1 RCT may not be feasible due to a limited patient population and other constraints. Alternative approaches can be considered, such as a hybrid randomized trial with an unequal randomization ratio (say, 3:1 experimental to control with proper augmentation of the control using concurrent external data) [15], a series of n-of-1 trials [16], a delayed-start design [17], platform trial designs [18], among others.

2.2.2 | Randomization Method

Randomization is the hallmark of clinical trials due to its potential for mitigating the risk of selection bias, ensuring comparability of treatment groups at baseline, and validating the use of statistical tests in the analysis. Various randomization methods can be used in practice [19], and the choice of a fit-for-purpose method for a given trial is not a trivial task [20].

2.2.3 | Statistical Analysis Method

The choice of a primary statistical analysis method is essential, as the study goal is to reliably detect clinically relevant treatment effects if they exist. A conventional method is the change-from-baseline analysis using two-sample *t*-test (2STT) or analysis of covariance (ANCOVA). However, these methods do not take into consideration the longitudinal nature of disease-modifying therapy trials. Some other methods may be more appropriate, such as mixed models with repeated measurements (MMRM) [21], or nonlinear mixed effects models (NLMEM) [22, 23].

2.2.4 | Statistical Analysis Approach

There are two approaches to the analysis of clinical trial data: a conventional population-based approach, which assumes random sampling from a target population, and a randomization-based approach, where experimental randomization forms the basis for statistical inference [19, 24]. While the two approaches yield similar *p*-values when the population model assumptions (normal random sampling) are satisfied, the randomization-based approach can be more robust when the assumptions are violated [25, 26]. Randomization-based tests have traditionally been applied in a very simplistic way, for example, using linear rank statistics; however, some recent work shows applications of randomization tests with longitudinal, survival, and generalized linear models [27]. To the best of our knowledge, a combination of PMx model-based analysis with randomization tests has not been investigated in the literature yet.

2.3 | Simulations

Once the clinical scenarios have been specified, they must be evaluated through Monte Carlo simulations. For each scenario, a clinical trial would be simulated repeatedly to evaluate the statistical properties of the considered options. Given rather complex PMx model-informed data generating mechanisms and PMx model-based analyses, a very powerful simulation engine must be in place.

2.4 | Performance Metrics

The performance metrics, also known as "design operating characteristics," are quantitative summaries of simulated distributions of the considered clinical scenarios. We distinguish three categories of performance metrics: statistical, logistical, and ethical. Statistical characteristics may include the type I error rate and power of statistical tests, the measures of estimation precision such as bias, variance, and mean squared error of estimators, etc. Logistical characteristics may include measures of operational efficiency such as time to complete the trial, total study cost, etc. Ethical characteristics may include some measures of benefit to the study participants such as the total number of successful treatment outcomes in the trial, average fraction of time on placebo treatment, etc.

2.5 | Decision Making

The final aim of the evaluation is to identify the best "trial setting" using a head-to-head comparison of the considered options across various experimental scenarios in terms of selected performance metrics. Visual displays of different metrics and their tradeoffs can help stakeholders make an informed decision and select the "optimal" available option for its implementation in practice. Ideally, a full report with all relevant documentation to ensure reproducible results should be supplied. Some best practices for projects that involve modeling and simulation can be useful in this regard [28, 29].

3 | Results

To illustrate the utility of our proposed CSE-PMx framework, we consider the problem of optimizing the design and analysis strategy of a hypothetical 1:1 RCT in ARSACS to compare the effects of an experimental treatment (E) versus control (C). We assume the total sample size is 100 patients (50 per arm) and the study duration for each participant is 2 years (24 months), with one visit every 6 months. We use a recently developed ARSACS disease progression model [30], which relates the total SARA (Scale for Assessment and Rating of Ataxia) score [31] to the time since onset (TSO) of the disease. The SARA scale has eight items, evaluated by a clinical assessment of gait, stance, speech, sitting, fine motor, and leg movements, and the total SARA score ranges from 0 (no ataxia) to 40 (most severe ataxia).

Following the Rubin Causal Model [32], we used a potential outcomes framework to simulate two possible trajectories of the total SARA score for each participant ($i = 1, \ldots, 100$): one assuming they received the experimental treatment ($Y_i^E(t)$), and the other one assuming they received the control ($Y_i^C(t)$). In reality, only one trajectory is observed, depending on the participant's

21638306, 2025, 10, Downloaded from https://ascpt.onlinelibrary.wiley.com/doi/10.1002/psp4.70082 by Uppsala University Library, Wiley Online Library on [30/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

treatment assignment. Specifically, we observe $Y_i^E(t)$ if the participant is assigned to the experimental group ($\xi_i = 1$), or $Y_i^C(t)$ if assigned to the control group ($\xi_i = 0$). The treatment effect at any given time point is then estimated by comparing the observed outcome with the unobserved (counterfactual) outcome under the alternative treatment.

To simulate $Y_i^C(t)$, we assume the natural course of disease progression. Patients with the history of ARSACS for not more than 30 years are deemed eligible for the trial. At baseline visit, the ith participant's time since onset of the disease (in months) is obtained as TSO_i. The participant's disease category at baseline is classified as "Early", if $0 < TSO_i < 120$; "Moderate", if $120 \le TSO_i < 240$; or "Late", if $240 \le TSO_i \le 360$. Using the disease progression model [30, 33], the longitudinal outcomes are simulated as

$$Y_i^C(t) = \delta_i + \frac{\gamma_i}{1 + \exp(\beta_i - \alpha_i t)}, \quad t \in \{t_{i0}, t_{i1}, t_{i2}, t_{i3}, t_{i4}\} \tag{1}$$

where, the individual parameters $(\delta_i, \gamma_i, \beta_i, \alpha_i)$ are assumed to follow independent log-normal distributions with typical values and uncertainties obtained from Hendrickx et al. (2024) [30]. The five observation time points include $t_{i0} = TSO_i$ (baseline), $t_{i1} = TSO_i + 6$, $t_{i2} = TSO_i + 12$, $t_{i3} = TSO_i + 18$, and $t_{i4} = TSO_i + 24$.

To simulate $Y_i^E(t)$, we assume that the experimental treatment has a disease modifying effect and slows down the increase in total SARA score. The longitudinal outcomes of the ith individual under treatment are simulated as

$$Y_i^E(t) = \delta_i + \frac{\gamma_i}{1 + \exp\left(\beta_i - \alpha_i TSO_i - \alpha_i (1 - DE)\left(t - TSO_i\right)\right)}, \quad t \in \{t_{i0}, t_{i1}, t_{i2}, t_{i3}, t_{i4}\} \tag{2}$$

where, the parameters $(\delta_i, \gamma_i, \beta_i, \alpha_i)$ and the sampling time points $\{t_{i0}, t_{i1}, t_{i2}, t_{i3}, t_{i4}\}$ are the same as in Equation (1), but an additional parameter, "Drug Effect", $DE \in [0, 1]$ determines the degree of inhibition of disease progression due to experimental treatment. For instance, $DE = 0$ implies no effect at all; $DE = 1$ implies full inhibition of disease progression; and $0 < DE < 1$ means partial inhibition of disease progression.

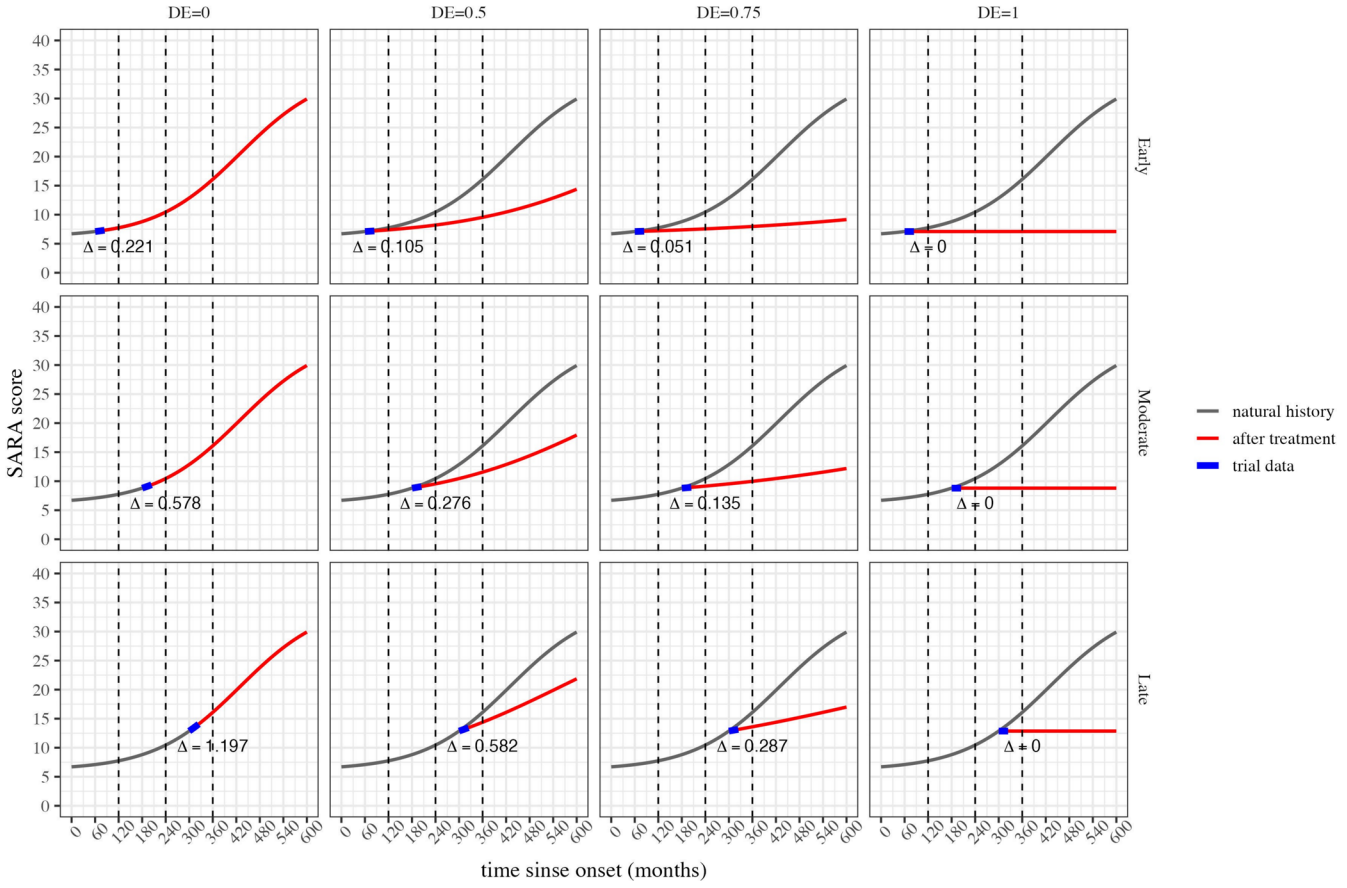

FIGURE 2 | A simulated example of an individual patient's ARSACS disease progression measured by the total SARA score over time. It is assumed that the disease spans 50 years (600 months). The gray curve represents the natural disease progression without treatment. The red curves represent treatment-induced changes in disease trajectory, which depend on the true values of drug effect (DE) and time since onset (TSO). DE = 0 means no effect; DE = 0.5 or 0.75 means 50% or 75% inhibition of disease progression; DE = 1 (ideal case) means 100% inhibition, that is, treatment arrests disease progression. For simplicity, it is assumed that treatment effects happen immediately after treatment administration. The earlier the efficacious treatment (with DE > 0) is given, the better is its potential to modify the course of the disease by maintaining smaller values of total SARA score over time. The blue-colored trajectories correspond to the true changes in total SARA score during the 2-year period of patient's participation in a clinical trial. The change from baseline to 24 months in total SARA score (Δ) depends on the value of DE and TSO at which the patient is administered the treatment.

21638306, 2025, 10, Downloaded from https://ascpt.onlinelibrary.wiley.com/doi/10.1002/psp4.70082 by Uppsala University Library, Wiley Online Library on [30/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Figure 2 shows a simulated example of the disease progression, expressed by the total SARA score over time, for an individual patient with ARSACS. It is assumed that the disease spans 50 years (600 months). The gray curve represents the natural course of the disease without treatment. Depending on the time of intervention (Early; Moderate; or Late) and the true treatment effect (DE = 0; 0.5; 0.75; or 1), the trajectory of the total SARA score may change. The red curves represent treatment-induced changes in disease trajectory, assuming for simplicity that treatment effects happen instantaneously after treatment administration. The earlier an efficacious treatment (with DE > 0) is given, the better is its disease-modifying potential as quantified by smaller values of total SARA score over time. The blue trajectories correspond to the true changes in total SARA score during the 2-year period of patient's participation in a clinical trial. The change from baseline to 24 months in total SARA score (Δ) depends on the value of DE and TSO at which the patient receives the treatment.

Using $Y_i^E(t)$ and $Y_i^C(t)$, the observed trajectory of the total SARA score for the ith participant is simulated as

$$Y_i(t) = Y_i^E(t)\xi_i + Y_i^C(t)\left(1-\xi_i\right) + \varepsilon_i(t) \quad (3)$$

where, $\varepsilon_i(t) \sim N\left(0, \sigma^2\right)$ are independent, identically distributed measurement error terms. In all simulations, we set $\sigma = 0.5$.

In clinical trials it is frequently assumed that study participants represent a random sample from the target population of patients. However, this assumption may be violated; for instance, there may be some hidden covariates correlated with the primary outcome that can negatively impact the type I error/power of a statistical test. To account for this possibility, we consider a scenario of "time trend" in the data, where patients with larger values of TSO are enrolled earlier in the trial. Such a phenomenon is not uncommon in clinical research—patients who are most severely impacted by the disease may wish to participate sooner and/or may be prioritized for inclusion in a study by clinical research sites. For this purpose, we generate a random sample (TSO_1, ..., TSO_{100}) from the discrete set $\{1, \ldots, 360\}$ and order the values from smallest to largest, such that the ordered sample corresponds to the order of patient entry in the trial.

Table 1 summarizes the simulation settings for our example. The considered assumptions for the data generating mechanism (Clinical scenarios: Assumptions) involve combinations of the four choices of the true drug effect (DE = 0; 0.5; 0.75; 1) and two possibilities of the presence of time trend (Yes or No). The statistical design and analysis strategy (Clinical scenarios: Options) involve combinations of two randomization methods (random allocation rule, Rand; or truncated binomial design, TBD), four statistical analysis methods (two-sample *t*-test, 2STT; analysis of covariance, ANCOVA; mixed model with repeated measurements, MMRM; or nonlinear mixed effects model, NLMEM); and two statistical analysis approaches (population-based or randomization-based).

A total of 128 scenarios ($= 4 \times 2 \times 2 \times 4 \times 2$) were explored. For each scenario, a 1:1 RCT with 100 participants was simulated 2500 times. The operating characteristics included the type I error rate under the "null" scenario (DE = 0), and statistical power to reject the null hypothesis of no treatment effect when the true effect was present (DE = 0.5; 0.75; 1).

Figure 3 summarizes the simulation results. Under the null scenario (DE = 0), most of the methods maintained type I error rate at the nominal level of 5% except for population-based 2STT with TBD, and population-based MMRM with either Rand or

TABLE 1 | Simulation settings for a 1:1 RCT with $N = 100$ ARSACS patients (50 per arm).

Data generating mechanism	Drug effect (DE)	{0, 0.5, 0.75, 1}
	Time trend determined by TSO	{No, Yes}
Statistical design and analysis strategy	Randomization design	{Rand, TBD}
	Statistical analysis method	{two-sample *t*-test, ANCOVA, MMRM, NLMEM}
	Statistical analysis approach	{Population-based, Randomization-based}

Note: Individual outcomes (total SARA score over time) with or without drug effect are generated using disease progression model [30, 32]. Study duration for each participant is 2 years (24 months), with one visit every 6 months; no missing data is assumed; measurement error standard deviation $\sigma = 0.5$. Four choices for the treatment effect: DE = 0 (no effect); DE = 0.5; 0.75 (partial inhibition of disease progression, by 50% and 75%, respectively); and DE = 1 (idealistic scenario, full inhibition of disease progression). Two choices for the presence of time trend in the data: "No" (we have a random sample of patients with TSO in the range 0–360 months); and "Yes" (time trend is present: patients with larger values of TSO are enrolled earlier in the trial). Two choices for the randomization design: random allocation rule (Rand), and truncated binomial design (TBD). Both randomization designs ensure equal final allocation (50 per arm) but have different probabilistic structures and different statistical properties (Rosenberger and Lachin, 2015). For Rand, one can think of an urn model, where an urn initially contains 50 balls of each type E and C. When a new patient enters the trial, a ball is drawn at random, without replacement, and the patient is assigned to treatment accordingly. The procedure is repeated until 100 patients are randomized into the study. For TBD, patients are randomized sequentially to E or C with probability 0.5, until one of the groups receives its target quota of 50 patients, after which the remaining patient are assigned (deterministically) to the opposite treatment. For Rand, all sequences of treatment assignments are equally likely. For TBD, some treatment sequences are more likely than others. Four choices for the statistical analysis method: (i) two-sample *t*-test (2STT) on the change from baseline to Month 24 in the total SARA score; (ii) Analysis of covariance (ANCOVA) of the change from baseline to Month 24 in the total SARA score, adjusting for the baseline values of total SARA score and TSO; (iii) Mixed model with repeated measurements (MMRM) using total SARA score post-baseline (Months 6, 12, 18, or 24) as the outcome, and fixed, categorical effects of treatment, visit, and treatment × visit interactions, as well as the continuous, fixed covariates of baseline (total SARA) score, and baseline score × visit interaction. An Unstructured (UN) covariance structure was used to model the within-subject errors; and (iv) Nonlinear mixed effects model (NLMEM) of the total SARA score (Hendrickx et al., 2024, Hamdan et al. 2024). Two choices for the statistical analysis approach: (i) Population-based approach for which the statistical analysis is carried out conditionally on the randomization design, and (ii) Randomization-based approach based on residuals (Parhat et al., 2024): for the chosen analysis method, a statistical model is fit under the null hypothesis of no treatment effect; the residuals (observed minus expected outcomes at the end of the study, i.e., Month 24) are calculated, and the strength of evidence against the null hypothesis is obtained by evaluating the tail probabilities of the re-randomization distribution of a linear rank test statistic based on residuals.

21638306, 2025, 10, Downloaded from https://ascpt.onlinelibrary.wiley.com/doi/10.1002/psp4.70082 by Uppsala University Library, Wiley Online Library on [30/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TBD. The fact that an unadjusted analysis such as 2STT following TBD is not robust to time trends due to potential deterministic allocations in the tail of TBD's randomization sequence is a well-known phenomenon [20, 34]. One risk mitigation strategy is to apply a randomization-based test following TBD or use ANCOVA adjusting for the covariate that caused time trend (in our case, TSO). The population-based MMRM substantially inflated type I error rate (up to 11%), a phenomenon attributed to model misspecification, as the linear MMRM analysis ignored the true nonlinear data generating process, leading to non-random residuals. To address the described deficiency of population-based MMRM, one can use randomization-based MMRM. From Figure 3, such an approach, maintained type I error rate (the lower limits of the 95% confidence intervals for the true type I error rate were < 5%). As for NLMEM, it resulted in valid type I error rate even when time trend was present, because data analysis used the correctly specified nonlinear model.

Regarding statistical power, several important observations can be made from Figure 3. First, although population-based MMRM demonstrated higher simulated power than population-based 2STT or ANCOVA, this apparent increase was misleading, as it was achieved at the expense of an inflated type I error rate. Table 2 reports the *corrected* power at DE = 0.5 for the population-based analysis methods that initially showed inflated type I error rates. The corrected power for MMRM ranged

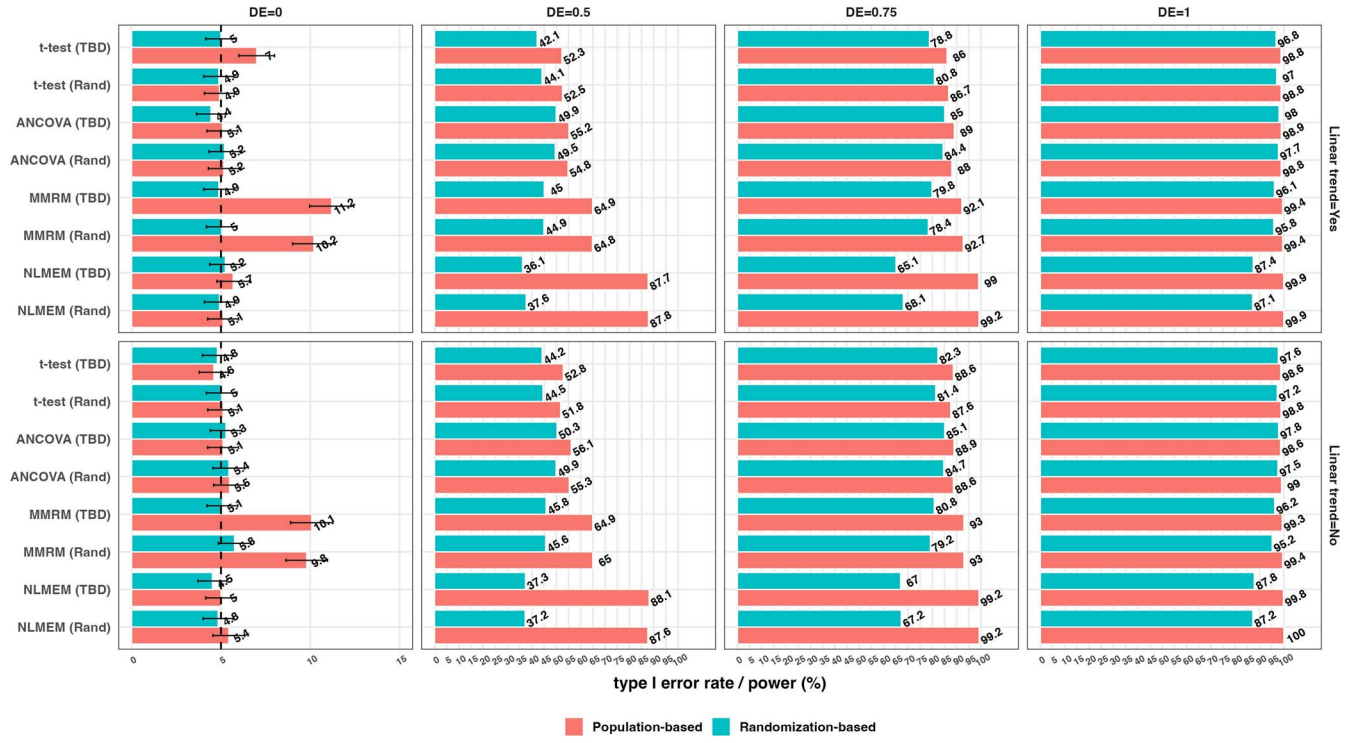

FIGURE 3 | Simulated type I error rate (for DE = 0) and power (for DE = 0.5, 0.75, 1) of 16 different statistical design/analysis strategies (options) under eight experimental settings defined by a combination of drug effect (DE) and presence of a time trend (Yes/No). For each scenario, a 1:1 RCT with 100 participants was simulated 2500 times.

TABLE 2 | Corrected power (at true DE = 0.5) that corresponds to the nominal type I error rate of 5% for the population-based analysis methods that had inflated type I error rate (cf. Figure 3).

Randomization design	Analysis method (population-based)	Time trend	Simulated T1ER (%)	Simulated power (%) at DE = 0.5	Corrected power (%) at the nominal T1ER = 5%
TBD	2STT	Yes	7.0	52.3	46.7
Rand	MMRM	Yes	10.2	64.8	52.7
TBD	MMRM	Yes	11.2	64.9	49.2
Rand	MMRM	No	9.8	65.0	51.2
TBD	MMRM	No	10.1	64.9	52.5

Note: To apply the correction, we simulated the distribution of each method's test statistic T under the null hypothesis H_0: DE = 0 and identified the 2.5th and 97.5th percentiles—denoted $t^*_{2.5}$ and $t^*_{97.5}$, respectively. These empirical critical values were then used to define a revised decision rule for simulations under the alternative hypothesis (DE = 0.5): reject H_0, if $T < t^*_{2.5}$ or $T > t^*_{97.5}$.

21638306, 2025, 10, Downloaded from https://ascpt.onlinelibrary.wiley.com/doi/10.1002/psp4.70082 by Uppsala University Library, Wiley Online Library on [30/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

from 49% to 53%, which was lower than that of ANCOVA (55%), a method that maintained a nominal type I error (see Figure 3).

Second, among the population-based analysis methods, NLMEM was clearly superior over other methods, irrespective of the implemented randomization (Rand or TBD). The observed difference was most pronounced when DE = 0.5—the power of NLMEM was 88%, whereas for other methods the (corrected) power was in the range 47%–55%. For DE = 0.75 or 1, NLMEM was still most powerful, but the absolute difference was less because the other methods also had high power. At the same time, randomization-based analysis with NLMEM was less powerful than randomization-based analysis with 2STT, ANCOVA, or MMRM.

Third, ANCOVA consistently demonstrated greater power than 2STT, regardless of whether the analysis was population-based or randomization-based, and irrespective of the randomization method used. This highlights the benefit of covariate adjustment in ANCOVA—in our case, adjusting for two key baseline covariates: TSO and baseline total SARA score.

Fourth, when comparing population-based and randomization-based approaches for a given analysis method (2STT, ANCOVA, MMRM, or NLMEM) and a given randomization method (Rand or TBD), population-based approaches were clearly more powerful. This is expected, as population-based approaches rely on parametric models, while randomization-based methods use nonparametric linear rank tests, which tend to be less efficient under correctly specified models.

Overall, in our example, the most effective strategy—both valid and powerful—is the population-based NLMEM, implemented with either Rand or TBD. This assumes that clinical investigators and stakeholders are comfortable with the nonlinear model and can appropriately interpret the estimated treatment effect. Alternatively, for those preferring a more traditional change-from-baseline analysis of the total SARA score, ANCOVA (whether population-based or randomization-based, and implemented with either Rand or TBD) offers a reasonable and interpretable option. In contrast, population-based MMRM may increase the risk of false positives, while randomization-based MMRM provides valid but potentially less powerful results.

4 | Summary and Discussion

The main contribution of the current paper is two-fold: (1) it proposes a PMx-informed clinical scenario evaluation framework (CSE-PMx) for comparing various design and analysis strategies for a disease-modifying clinical trial of a given RND, including generation and provision of a simulation engine; and (2) it provides an illustrative example of applying CSE-PMx in the context of a hypothetical randomized clinical trial in RND, exemplified for the ultra-rare disease patients with ARSACS. Our proposal extends the existing CSE framework [13] by including a nonlinear disease progression model for generating longitudinal outcomes over time, considering different randomization methods, and including an option for implementing randomization-based tests that can protect against spurious significant results caused by temporal trends.

Our considered example was a relatively small randomized controlled trial (a total of 100 patients, 50 per arm), with a relatively short study duration of 24 months, with clinical visits every 6 months. The study objective was to compare the disease-modifying effect of an experimental treatment versus control, with respect to a 24-month change in total SARA score. For the given setup, our assessment of different options spanned three dimensions: the choice of a randomization method, statistical analysis method, and statistical analysis approach (population- or randomization-based). Our simulation evidence suggests that a NLMEM with a population-based likelihood ratio test analysis is valid, robust to possible time trends (caused by a known variable, captured as a baseline covariate), and more powerful than some conventional methods such as two-sample *t*-test, ANCOVA, or MMRM. Unlike the latter methods, NLMEM uses observations at intermediate time points and accounts for the structure of the time-response relationship of the primary outcome for every individual as informed by the disease progression model from historical data. In our simulations, an excellent performance of NLMEM can be attributed to the fact that the analysis model was fully aligned with the data-generating mechanism. In practice, however, the working model may be misspecified in various ways. While this study did not explore the impact of model misspecification, it remains an important area for future research.

It is also worth noting that NLMEM and two-sample *t*-test/ANCOVA/MMRM consider different estimands: NLMEM is aimed at estimating/testing the significance of the degree of inhibition of disease progression due to the experimental treatment in comparison to the control, whereas for the other methods, the target estimand is the mean group difference in change from baseline to the end of study (in our case, 24 months) of the primary outcome. Clearly, these estimands have different clinical interpretations and address somewhat different research questions.

We would like to highlight some limitations of our current work and outline some opportunities for future research. In our simulation study, we only explored a parallel group, 1:1 RCT design. However, other designs may be also considered, such as a series of n-of-1 trials [16], a randomized delayed start design [17], a crossover design, etc. Adding "design" dimension to expand the set of simulations merits investigation [35]. In our study, we focused on the total SARA score as the primary outcome. This score is a quasi-continuous measure representing a sum of item responses (correlated ordinal outcomes). Alternatively, one can consider item response theory (IRT)-informed model [33, 36, 37] to simulate individual outcomes accounting for relevant covariates and uncertainties and perform an IRT model-based analysis to test the group differences. Such an approach may be even more powerful than NLMEM of the total SARA score. Another important aspect that has not been explored in our simulations is the impact of missing data. In a clinical trial, some participants are likely to drop out, there may be missing clinical visits, etc. One may add the dimension "pattern and percentage of missing data" to examine the robustness of different design and analysis strategies to these phenomena. Finally, it is still an open question how to construct a powerful randomization test based on NLMEM or MMRM. Our explored approach using a linear rank test of residuals from NLMEM or MMRM fitted under the null hypothesis of no treatment effect. Such an approach is valid

21638306, 2025, 10, Downloaded from https://ascpt.onlinelibrary.wiley.com/doi/10.1002/psp4.70082 by Uppsala University Library, Wiley Online Library on [30/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

but not efficient. Developing a better, computationally feasible method is an important open problem.

Author Contributions

O.S., Y.R., R.-D.H., N.H., E.C., F.M., N.H., M.O.K., A.C.H., A.H., X.C., R.S., and M.S. wrote the manuscript; O.S., Y.R., R.-D.H., N.H., E.C., F.M., N.H., M.O.K., A.C.H., and X.C. designed the research; Y.R. and O.S. performed the research; Y.R. and O.S. analyzed the data.

Acknowledgments

We thank Dr. Dr. Andreas Traschütz, Center of Neurology, University of Tübingen for his input in discussions on data that were included in this manuscript. We thank the PROSPAX consortium for providing the data for earlier studies, which helped to inform the analyses performed and presented in the current manuscript. We would also like to thank the two anonymous reviewers for their insightful comments and constructive suggestions on the manuscript.

Disclosure

The views expressed in this article are the personal views of the author(s) and may not be construed or quoted as being made on behalf of or reflecting the position of the organizations with which the author(s) e employed/affiliated. As Editor-in-Chief of *CPT: Pharmacometrics & Systems Pharmacology*, France Mentré was not involved in the review or decision process for this paper. As an Associate Editor for *CPT: Pharmacometrics & Systems Pharmacology*, Andrew Hooker was not involved in the review or decision process for this paper.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The computer code used to generate the simulation data and perform the statistical analyses presented in this manuscript is publicly available through the first author's Github repository https://github.com/yevgenryeznik/cse-pmx.

References

1. The Lancet Neurology, "Rare Diseases: Maintaining Momentum," *Lancet Neurology* 21, no. 3 (2022): 203, https://doi.org/10.1016/S1474-4422(22)00046-1.

2. European Medicines Agency, "Guideline on Clinical Trials in Small Populations," 2006, https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-clinical-trials-small-populations_en.pdf.

3. R. D. Hilgers, M. Bogdan, C. F. Burman, et al., "Lessons Learned From IDeAl - 33 Recommendations From the IDeAl-Net About Design and Analysis of Small Population Clinical Trials," *Orphanet Journal of Rare Diseases* 13, no. 1 (2018): 77, https://doi.org/10.1186/s13023-018-0820-8.

4. R. D. Hilgers, F. König, G. Molenberghs, and S. Senn, "Design and Analysis of Clinical Trials for Small Rare Disease Populations," *Journal of Rare Diseases Research & Treatment* 1, no. 3 (2016): 53–60, https://doi.org/10.29245/2572-9411/2016/3.1054.

5. R. D. Hilgers, K. Roes, N. Stallard, IDeAl, Asterix, and InSPiRe, "Directions for New Developments on Statistical Design and Analysis of Small Population Group Trials," *Orphanet Journal of Rare Diseases* 11 (2016): 78, https://doi.org/10.1186/s13023-016-0464-5.

6. T. Friede, M. Posch, S. Zohar, et al., "Recent Advances in Methodology for Clinical Trials in Small Populations: The InSPiRe Project," *Orphanet Journal of Rare Diseases* 13 (2018): 186, https://doi.org/10.1186/s13023-018-0919-y.

7. EFPIA MID3 Workgroup, S. F. Marshall, R. Burghaus, et al., "Good Practices in Model-Informed Drug Discovery and Development: Practice, Application, and Documentation," *CPT Pharmacometrics & Systems Pharmacology* 5, no. 3 (2016): 93–122, https://doi.org/10.1002/psp4.12049.

8. S. Marshall, R. Madabushi, E. Manolis, et al., "Model-Informed Drug Discovery and Development: Current Industry Good Practice and Regulatory Expectations and Future Perspectives," *CPT: Pharmacometrics & Systems Pharmacology* 8, no. 2 (2019): 87–96, https://doi.org/10.1002/psp4.12372.

9. G. Koch, M. Pfister, I. Daunhawer, M. Wilbaux, S. Wellmann, and J. E. Vogt, "Pharmacometrics and Machine Learning Partner to Advance Clinical Data Analysis," *Clinical Pharmacology & Therapeutics* 107, no. 4 (2020): 926–933, https://doi.org/10.1002/cpt.1774.

10. Y. Ryeznik, O. Sverdlov, E. M. Svensson, G. Montepiedra, A. C. Hooker, and W. K. Wong, "Pharmacometrics Meets Statistics—A Synergy for Modern Drug Development," *CPT: Pharmacometrics & Systems Pharmacology* 10, no. 10 (2021): 1134–1149, https://doi.org/10.1002/psp4.12696.

11. S. A. G. Visser, J. D. Norton, S. Marshall, and M. O'Kelly, "Common Best Practice in Modeling and Simulation Across Quantitative Disciplines: A Comparison of Independently Emerging Proposals," *Statistics in Biopharmaceutical Research* 10, no. 2 (2018): 72–75, https://doi.org/10.1080/19466315.2017.1385520.

12. N. Benda, M. Branson, W. Maurer, and T. Friede, "Aspects of Modernizing Drug Development Using Clinical Scenario Planning and Evaluation," *Therapeutic Innovation & Regulatory Science* 44, no. 3 (2010): 299–315, https://doi.org/10.1177/009286151004400312.

13. T. Friede, R. Nicholas, N. Stallard, et al., "Refinement of the Clinical Scenario Evaluation Framework for Assessment of Competing Development Strategies With an Application to Multiple Sclerosis," *Therapeutic Innovation & Regulatory Science* 44, no. 6 (2010): 713–718, https://doi.org/10.1177/009286151004400607.

14. R. D. Hilgers, D. Uschner, W. F. Rosenberger, and N. Heussen, "ERDO – A Framework to Select an Appropriate Randomization Procedure for Clinical Trials," *BMC Medical Research Methodology* 17 (2017): 159, https://doi.org/10.1186/s12874-017-0428-z.

15. S. Ventz, S. Khozin, B. Louv, et al., "The Design and Evaluation of Hybrid Controlled Trials That Leverage External Data and Randomization," *Nature Communications* 13, no. 1 (2022): 5783, https://doi.org/10.1038/s41467-022-33192-1.

16. E. O. Lillie, B. Patay, J. Diamant, B. Issell, E. J. Topol, and N. J. Schork, "The n-of-1 Clinical Trial: The Ultimate Strategy for Individualizing Medicine?," *Personalized Medicine* 8, no. 2 (2011): 161–173, https://doi.org/10.2217/pme.11.7.

17. R. B. D'Agostino, "The Delayed-Start Study Design," *New England Journal of Medicine* 36, no. 113 (2019): 1304–1306.

18. Adaptive Platform Trials Coalition, "Adaptive Platform Trials: Definition, Design, Conduct and Reporting Considerations," *Nature Reviews Drug Discovery* 18, no. 10 (2019): 797–807, https://doi.org/10.1038/s41573-019-0034-3.

19. W. F. Rosenberger and J. Lachin, *Randomization in Clinical Trials: Theory and Practice*, 2nd ed. (Wiley, 2015).

20. V. W. Berger, L. J. Bour, K. Carter, et al., "A Roadmap to Using Randomization in Clinical Trials," *BMC Medical Research Methodology* 21, no. 1 (2021): 168, https://doi.org/10.1186/s12874-021-01303-z.

21. C. H. Mallinckrodt, P. W. Lane, D. Schnell, Y. Peng, and J. P. Mancuso, "Recommendations for the Primary Analysis of Continuous Endpoints in Longitudinal Clinical Trials," *Drug Information Journal* 42 (2008): 303–319, https://doi.org/10.1177/009286150804200402.

21638306, 2025, 10, Downloaded from https://ascpt.onlinelibrary.wiley.com/doi/10.1002/psp4.70082 by Uppsala University Library, Wiley Online Library on [30/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

22. L. B. Sheiner, B. Rosenberg, and K. L. Melmon, "Modelling of Individual Pharmacokinetics for Computer-Aided Drug Dosage," *Computers and Biomedical Research* 5 (1972): 441–459, https://doi.org/10.1016/0010-4809(72)90051-1.

23. M. L. Lindstrom and D. M. Bates, "Nonlinear Mixed Effects Models for Repeated Measures Data," *Biometrics* 46, no. 3 (1990): 673–687, https://doi.org/10.2307/2532087.

24. R. A. Fisher, "The Design of Experiments," 1935.

25. M. Proschan and L. E. Dodd, "Re-Randomization Tests in Clinical Trials," *Statistics in Medicine* 38, no. 12 (2019): 2292–2302, https://doi.org/10.1002/sim.8093.

26. W. F. Rosenberger, D. Uschner, and Y. Wang, "Randomization: The Forgotten Component of the Randomized Clinical Trial," *Statistics in Medicine* 38, no. 1 (2019): 1–12, https://doi.org/10.1002/sim.7901.

27. P. Parhat, W. F. Rosenberger, and G. Diao, "Conditional Monte Carlo Randomization Tests for Regression Models," *Statistics in Medicine* 33 (2014): 3078–3088, https://doi.org/10.1002/sim.6149.

28. M. O'Kelly, V. Anisimov, C. Campbell, and S. Hamilton, "Proposed Best Practice for Projects That Involve Modelling and Simulation," *Pharmaceutical Statistics* 16, no. 2 (2017): 107–113, https://doi.org/10.1002/pst.1789.

29. C. Mayer, I. Perevozskaya, S. Leonov, et al., "Simulation Practices for Adaptive Trial Designs in Drug and Device Development," *Statistics in Biopharmaceutical Research* 11, no. 4 (2019): 325–335, https://doi.org/10.1080/19466315.2018.1560359.

30. N. Hendrickx, F. Mentré, A. Traschütz, et al., "Prediction of Individual Disease Progression Including Parameter Uncertainty in Rare Neurodegenerative Diseases: The Example of Autosomal-Recessive Spastic Ataxia Charlevoix Saguenay (ARSACS)," *AAPS Journal* 26, no. 3 (2024): 57, https://doi.org/10.1208/s12248-024-00925-7.

31. T. Schmitz-Hübsch, S. T. du Montcel, L. Baliko, et al., "Scale for the Assessment and Rating of Ataxia: Development of a New Clinical Scale," *Neurology* 66, no. 11 (2006): 1717–1720, https://doi.org/10.1212/01.wnl.0000219042.60538.92.

32. D. B. Rubin, "Causal Inference Using Potential Outcomes: Design, Modeling, Decisions," *Journal of the American Statistical Association* 100, no. 469 (2005): 322–331, https://www.jstor.org/stable/27590541.

33. A. Hamdan, N. Hendrickx, A. C. Hooker, et al., "Longitudinal Analysis of Natural History Progression of Rare and Ultra-Rare Cerebellar Ataxias Using Item Response Theory," *Clinical Pharmacology and Therapeutics* 116, no. 6 (2024): 1593–1605, https://doi.org/10.1002/cpt.3466.

34. G. K. Rosenkranz, "The Impact of Randomization on the Analysis of Clinical Trials," *Statistics in Medicine* 30, no. 30 (2011): 3475–3487, https://doi.org/10.1002/sim.4376.

35. N. Hendrickx, F. Mentré, A. Hamdan, et al., "Comparing Randomized Trial Designs to Estimate Treatment Effect in Rare Diseases With Longitudinal Models: A Simulation Study Showcased by Autosomal Recessive Cerebellar Ataxias Using the SARA Score," *BMC Medical Research Methodology* (2025).

36. S. Ueckert, E. L. Plan, K. Ito, M. O. Karlsson, B. Corrigan, and A. C. Hooker, "Improved Utilization of ADAS-Cog Assessment Data Through Item Response Theory Based Pharmacometric Modeling," *Pharmaceutical Research* 31, no. 8 (2014): 2152–2165, https://doi.org/10.1007/s11095-014-1315-5.

37. G. J. Wellhagen, S. Ueckert, M. C. Kjellsson, and M. O. Karlsson, "An Item Response Theory–Informed Strategy to Model Total Score Data From Composite Scales," *AAPS Journal* 23 (2021): 45, https://doi.org/10.1208/s12248-021-00555-3.

21638306, 2025, 10, Downloaded from https://ascpt.onlinelibrary.wiley.com/doi/10.1002/psp4.70082 by Uppsala University Library, Wiley Online Library on [30/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License