Check for updates

Clinical and Translational Science

REVIEW OPEN ACCESS

Supply and Demand in the Mathematics of Rare Disease Drug Development: Why Choosing the Right Model Is Crucial

Marshall L. Summar[1] | Janet Woodcock[2]

[1]George Washington University, CEO, Uncommon Cures, Washington DC, USA | [2]Center for Drug Evaluation and Research, U.S. Food and Drug Administration, Silver Spring, Maryland, USA

Correspondence: Marshall L. Summar (msummar@uncommoncures.com)

Received: 19 February 2026 | **Revised:** 18 March 2026 | **Accepted:** 20 March 2026

Keywords: clinical trial design | natural history controls | patient-as-own-control | rare diseases | sample size | statistical power

ABSTRACT

Clinical trials for rare diseases face a fundamental mathematical challenge that conventional randomized controlled trial (RCT) designs cannot overcome. With approximately 95% of the estimated 10,000–16,000 rare diseases lacking approved therapies, and drug development programs failing at rates exceeding 75% in non-oncology indications, the field confronts a stark reality: Traditional trial designs demand patient numbers that simply do not exist. This perspective article examines the critical mismatch between the statistical requirements of different trial designs (the "demand") and the actual patient populations available for study (the "supply"). We demonstrate mathematically that alternative trial designs—particularly patient-as-own-control and natural history comparator models—can reduce required sample sizes by 5- to 20-fold while maintaining statistical rigor. We further point out that a substantial proportion of rare disease trial failures stem not from therapeutic inefficacy but from recruitment and retention challenges inherent to underpowered RCT designs—challenges that are directly addressable through appropriately matched trial design. Given that most rare disease development programs receive only one opportunity to demonstrate efficacy, the continued application of inappropriate statistical models represents both a scientific failure and an ethical and economic challenge to the rare disease community. We propose that regulatory agencies formalize acceptance of alternative trial designs for rare diseases, supported by explicit mathematical frameworks that transparently account for genetic heterogeneity, pediatric populations, and the statistical efficiency gains achieved through within-subject correlation.

1 | Introduction

The randomized controlled trial (RCT) has been the cornerstone of evidence-based medicine since Sir Austin Bradford Hill's seminal 1948 streptomycin trial for tuberculosis [1]. This design—characterized by random assignment to treatment or control groups, blinding, and concurrent controls—was engineered to eliminate bias, control confounding variables, and establish causality with high statistical confidence. The concept of controlled comparison in medical research predates the modern RCT by more than 150 years. In 1784, a royal commission appointed by King Louis XVI—including Benjamin Franklin and Antoine Lavoisier—conducted what is often considered the first blinded controlled trial, designed to evaluate Franz Mesmer's claims of "animal magnetism." Subjects were blindfolded and exposed to either genuine or sham magnetic treatments; when they could not distinguish between the two, the commission concluded that the purported effects were attributable to imagination rather than any physical force. This foundational experiment established the principle that therapeutic claims require controlled comparison to separate real effects from expectation and bias [2].

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.

© 2026 The Author(s). *Clinical and Translational Science* published by Wiley Periodicals LLC on behalf of American Society for Clinical Pharmacology and Therapeutics.

For common diseases where patient populations are abundant and effect sizes often substantial, RCTs have proven remarkably successful, earning their designation as the "gold standard" for regulatory approval worldwide. However, when applied to rare diseases, the RCT's fundamental requirements become insurmountable obstacles.

Rare diseases are defined variably across jurisdictions, but all definitions share a common consequence: Most of the included patient populations are too small to power conventional trials. In the United States, a rare disease affects fewer than 200,000 individuals; in the European Union, the prevalence must not exceed 5 per 10,000; in Japan, fewer than 50,000 patients nationwide. Yet even these seemingly generous thresholds belie the true challenge. Among the more than 10,000 identified rare diseases, the vast majority affect far fewer patients—with 80-90% classified as "extremely-rare" (prevalence < 1 per 100,000) and nearly 40% affecting fewer than 10 patients worldwide [3, 4].

The consequences of these small numbers are well documented. Analyses of clinical trial outcomes demonstrate that non-oncology orphan drug development programs fail at rates exceeding 75%, with a substantial proportion of these failures attributable to recruitment and retention difficulties—including inadequate statistical power due to inappropriate trial design—rather than true lack of therapeutic efficacy [5–8]. The precise breakdown of failure causes varies across analyses and has not been comprehensively characterized in rare disease-specific datasets, but recruitment and retention consistently emerge as leading contributors. This distinction is crucial: Many potentially effective therapies are abandoned not because they do not work, but because the trial design selected could not detect whether they worked. When a rare disease program fails due to inadequate statistical power, patients lose access to what might have been their only therapeutic option, and sponsors—often small biotechnology companies, patient organizations, or academic institutions—lack the resources to attempt a second trial.

1.1 | Three Alternative Trial Designs

This perspective examines three trial designs that offer statistically rigorous solutions for rare disease research. The double-blind placebo-controlled randomized controlled trial (DBPCRCT) serves as our comparator—the traditional design in which patients are randomly assigned to receive either active treatment or placebo, with neither patients nor investigators aware of assignments. The patient-as-own-control design uses each participant as their own baseline comparator, encompassing several subtypes: *Crossover trials*, in which patients receive both active treatment and placebo in randomized sequence with a washout period between; *run-in designs*, in which baseline measurements during a pre-treatment observation period are compared to post-treatment outcomes, particularly suited for irreversible interventions such as gene therapy; and within subject dose titration (also referred to N-of-1 trials, not to be confused with a treatment developed for just one patient) using treatment-withdrawal cycles to establish personal treatment response. The natural history (or external control) comparator design compares outcomes in treated patients to historical data from untreated individuals, including: *Matched historical controls*, in which treated patients are paired with historical patients based on demographic and clinical characteristics using propensity score matching; *registry-based comparisons*, which leverage established disease registries such as the Cystic Fibrosis Foundation Patient Registry to provide well-characterized control populations; and *sibling or familial controls*, particularly valuable in genetic diseases where affected but untreated family members provide biologically matched comparators.

Choosing among these three designs requires matching statistical and practical characteristics to the disease context. Patient-as-own-control designs are most appropriate when treatment effects are reversible (crossover) or when stable pre-treatment baselines can be established (run-in), disease progression is predictable, and endpoints are continuous and objective. Natural history comparator designs are preferred when interventions are irreversible (particularly gene therapies), when a well-characterized natural history dataset exists, or when crossover is not feasible due to persistent treatment effects. Natural history designs should be favored over crossover when carryover effects cannot be adequately addressed by a washout period, or when the disease trajectory is sufficiently well-characterized that historical controls provide reliable counterfactual estimates. Adaptive designs are appropriate when prior data are limited, and design parameters—sample size, endpoints, patient selection—may require modification based on interim evidence. Importantly, efficiency ratios between designs are constant across effect sizes: The fold-reduction from RCT to own-control remains approximately 19× regardless of whether $d = 0.25$ or $d = 1.0$, because all formulas scale as $1/d^2$. The absolute number of required patients varies substantially with effect size; however, making the choice of design particularly consequential for diseases with small available populations.

1.2 | The Challenge of Genetic Heterogeneity

Approximately 72% of rare diseases have identified genetic origins [9], yet even within a single genetic diagnosis, phenotypic expression varies enormously. Consider Duchenne muscular dystrophy (DMD): While all patients share dystrophin gene mutations, the specific mutation, modifier genes, and epigenetic factors produce wide variation in disease severity, age of onset, and progression rate. In a traditional RCT comparing treated vs. control groups, this inter-individual variation manifests as "noise" that obscures treatment signals, requiring larger sample sizes to achieve statistical significance.

The patient-as-own-control design elegantly addresses this heterogeneity by using each patient as their own baseline comparator. Because an individual's genetic background, modifier genes, and baseline disease state remain constant, comparing pre- and post-treatment measurements within the same patient eliminates inter-individual variation entirely. The relevant correlation (ρ) between repeated measurements on the same individual typically exceeds 0.85–0.95 for biological parameters, dramatically reducing the variance of the treatment effect estimate and, consequently, the sample size required to detect that effect.

17528062, 2026, 5, Downloaded from https://ascpt.onlinelibrary.wiley.com/doi/10.1111/cts.70551 by Readcube (Labtiva Inc.), Wiley Online Library on [28/10/2026]. 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

1.3 | The Pediatric Exclusion Problem

Compounding the challenge, 75% of rare diseases have pediatric onset [9, 10], yet children are systematically excluded from clinical trials due to ethical concerns, regulatory complexity, and recruitment difficulties. This exclusion has profound consequences. When trials are conducted exclusively in adults, investigators may be testing therapeutics at the least sensitive point in the disease trajectory—after years of accumulated damage have reduced the potential for therapeutic response. A gene therapy administered to an adult with a metabolic disorder may face irreversible organ damage that would not be present in a pediatric patient.

Furthermore, disease progression often differs fundamentally between pediatric and adult patients. Adult studies may demonstrate no effect on endpoints that would have shown dramatic improvement in children at disease onset. The within-subject designs we advocate can facilitate pediatric enrollment by reducing total patient requirements and by minimizing placebo exposure—a particularly important consideration for parents weighing their child's participation.

This perspective article presents a quantitative framework for understanding why conventional RCT designs fail in rare diseases and how alternative designs—specifically patient-as-own-control and natural history comparator models—offer mathematically rigorous solutions. We approach this challenge as a supply-and-demand problem: The "supply" represents available patients, while "demand" represents the sample sizes required by different statistical designs to detect treatment effects with adequate power (Figure 1).

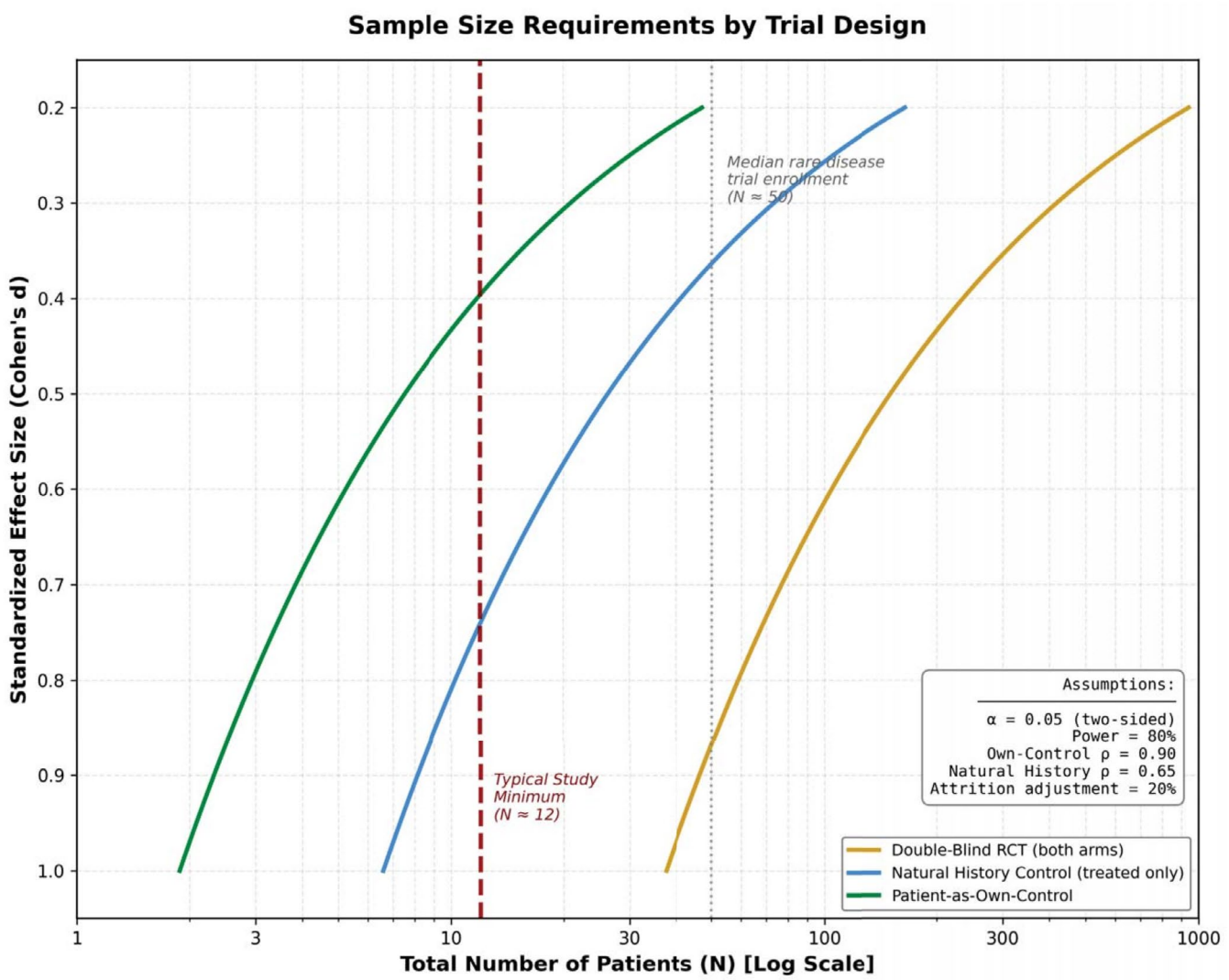


FIGURE 1 | Sample size requirements for three clinical trial designs as a function of standardized effect size (Cohen's d), calculated for 80% power at $\alpha = 0.05$ (two-tailed). The x-axis displays total required patients (N) on a logarithmic scale; the y-axis displays effect size ranging from 0.2 (small) to 1.0 (large). The gold curve represents the double-blind placebo-controlled RCT, calculated as $N = 37.63/d^2$. The blue curve represents the matched natural history comparator design with correlation $\rho = 0.65$, calculated as $N = 6.59/d^2$. The green curve represents the patient-as-own-control design with within-subject correlation $\rho = 0.90$, calculated as $N = 1.88/d^2$. All formulas include a 20% adjustment to account for anticipated patient attrition, protocol deviations, and missing data. The logarithmic scale emphasizes the dramatic efficiency differences at smaller sample sizes typical of rare disease research. Sample size requirements at specific effect sizes are provided in Table 3.

17528062, 2026, 5, Downloaded from https://ascpt.onlinelibrary.wiley.com/doi/10.1111/cts.70551 by Readcube (Labtiva Inc.), Wiley Online Library on [28/10/2026]. 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 | The Supply: Patient Availability in Rare Disease

The fundamental constraint in rare disease drug development is patient availability. While definitions vary globally, the common thread is scarcity. To quantify this challenge, we analyzed reported prevalence data from Orphanet, the world's largest database of rare diseases, focusing on the 3128 conditions with documented prevalence estimates [11, 12]. Table 1 may overestimate prevalence for newer diseases since there are few prevalence reports on disorders that are newly described for single or a handful of patients.

The data reveal a sobering reality: 95% of rare diseases (2,970 of 3,128) fall into the ultra-rare category with prevalence below 1 per 100,000. Within this group, 38% (1,188 diseases) are extremely rare—defined here as prevalence below 0.1 per 100,000—with fewer than 10 known cases worldwide. For the majority of rare diseases, the entire global patient population falls below the sample size required for a conventional RCT.

3 | The Demand: Statistical Requirements of Trial Designs

The sample size required to detect a given treatment effect depends fundamentally on the variance of the treatment effect estimate. Different trial designs achieve different levels of variance reduction, translating directly to different sample size requirements. Here we present the mathematical foundations of three major designs.

TABLE 1 | Patient Availability by Disease Rarity.

Category	Diseases	% of Total/ Cumulative %	U.S. Patients
Extremely Rare (<0.1/100,000) — subcategory of Ultra-rare			
Fewer than 10 worldwide	1,188	38%/38%	0–1
10–100 worldwide	891	28%/66%	0–5
100–1000 worldwide	594	19%/85%	4–43
1000–8000 worldwide	297	10%/95%	43–341
Rare (0.1–50/100,000)			
8,000–100,000 worldwide	125	4%/99%	341–4,260
100,000–4,000,000 worldwide	33	1%/100%	4,260–170,500

Note: Orphanet database; 3,128 diseases with documented prevalence [11, 12]. U.S. population: 341 million.

3.1 | Key Statistical Concepts for Non-Statisticians

Table 2.

3.2 | Adjustment for Real-World Conditions

The sample size formulas presented below represent theoretical minima assuming perfect data collection. In practice, clinical trials encounter inevitable complications that reduce the effective sample size: Patient attrition due to adverse events, disease progression, or withdrawal of consent; protocol deviations that render some data unusable; missing measurements from failed laboratory assays or missed visits; and patients who prove ineligible after enrollment. To account for these realities, we apply a 20% inflation factor to all calculated sample sizes—a correction widely adopted in clinical trial planning [13]. This adjustment is conservative for well-designed rare disease trials, which often achieve higher retention rates due to strong patient engagement and limited alternative treatment options, but provides a reasonable margin for regulatory acceptance. The adjusted formulas thus represent practical, defensible targets rather than theoretical ideals.

3.3 | Design 1: Double-Blind Placebo-Controlled RCT

In a standard 1:1 randomized design, treatment and control samples are independent. Throughout all three designs presented in this article, the effect size d is standardized by the pooled between-subject standard deviation, ensuring cross-design comparability and alignment with published effect size estimates from the literature. When two observations are statistically independent, the variance of their difference is additive [14]:

TABLE 2 | Variable Definitions.

Term	Definition
d (effect size)	Standardized treatment effect: $d = (\text{mean}_{\text{treatment}} - \text{mean}_{\text{control}})/\text{pooled SD}$. Values of 0.2, 0.5, and 0.8 represent small, medium, and large effects.
α (alpha)	Type I error rate; probability of false positive. Standard: $\alpha = 0.05$ (two-tailed).
β (beta)	Type II error rate; probability of false negative. Power $= 1 - \beta$.
Power $(1 - \beta)$	Probability of detecting a true effect. Standard: 80% (0.80).
$Z_{\alpha/2}$	Critical value for significance level. For $\alpha = 0.05$: $Z_{\alpha/2} = 1.96$.
Z_{β}	Critical value for power. For 80% power: $Z_{\beta} = 0.84$.
ρ (rho)	Correlation between paired measurements. Higher ρ reduces required sample size.
N	Total sample size required (treated patients + controls where applicable).

17528062, 2026, 5, Downloaded from https://ascpt.onlinelibrary.wiley.com/doi/10.1111/cts.70551 by Readcube (Labtiva Inc.), Wiley Online Library on [28/10/2026]. 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

$$\mathrm{Var}\left(Y_T - Y_C\right) = 2\sigma^2$$

The sample size per arm required to achieve power $(1-\beta)$ at significance level α is:

$$N_{RCT} = 4\left(Z_{\alpha/2} + Z_\beta\right)^2 / d^2$$

For $\alpha = 0.05$ and 80% power, the numerical substitution proceeds as follows: $Z_{\alpha/2} + Z_\beta = 1.96 + 0.84 = 2.80$; $(2.80)^2 = 7.84$; $4 \times 7.84 = 31.36$. Thus:

$$N_{RCT} = 31.36 / d^2 \text{ (base formula)}$$

With a 20% adjustment for attrition and practical considerations:

$$N_{RCT} = 37.63 / d^2 (\text{adjusted})$$

At an effect size of $d = 0.5$, this translates to approximately 151 total patients (76 per arm). Note that these formulas employ large-sample Z-based (normal) approximations, which are used throughout this article for cross-design planning and comparison. At very small sample sizes—as commonly encountered in ultra-rare diseases—exact t-distribution-based or simulation-based power calculations are preferable for actual trial planning. The 20% attrition adjustment partially compensates for approximation error at small N, but sponsors should perform exact calculations for regulatory submissions.

For rare disease trials, a 2:1 allocation ratio (twice as many treated patients as controls) is more commonly employed than 1:1, as it maximizes treated-patient access while preserving a concurrent control arm. Under 2:1 allocation, the total sample size formula becomes $N_RCT(2:1) = (9/4) \times (Z\alpha/2 + Z\beta)^2 \times (1 + 1/r) / d^2$, where $r = 2$, yielding $N_RCT(2:1) \approx 37.63 \times (3/2) / d^2 \approx 56.44 / d^2$ (adjusted). At $d = 0.5$, this requires approximately 226 total patients—modestly more than the 151 required under 1:1, but still requiring far more patients than either alternative design. This modest efficiency cost is reflected in Table 3. The efficiency argument for alternative designs thus holds even when the most patient-friendly standard RCT allocation is used.

3.4 | Design 2: Matched Natural History Comparator

When treated patients are matched to historical controls based on baseline characteristics, the measurements become correlated. If ρ represents the correlation between matched pairs, the variance of the difference score is reduced [15]:

$$\mathrm{Var}(D) = 2\sigma^2(1 - \rho)$$

The required sample size becomes:

$$N_{NH} = 2(1 - \rho)\left(Z_{\alpha/2} + Z_\beta\right)^2 / d^2$$

For illustrative purposes, we use $\rho \approx 0.65$ as an example planning value, consistent with the degree of covariate balance typically achievable through propensity score matching when accounting for measured confounders, temporal differences, and measurement variability [16, 17]. This value should not be adopted as a default; the correlation actually achieved depends critically on the quality of the natural history dataset, the number of matching variables, and the degree of unmeasured confounding in the specific disease context. Sponsors should derive ρ empirically from available natural history data wherever possible (see Table S1 for sensitivity of required sample size to ρ):

$$N_{NH} = 5.49 / d^2 (\text{base}) \rightarrow N_{NH} = 6.59 / d^2 (\text{adjusted})$$

Table S1 (Sensitivity Analysis). Required sample size for the natural history comparator design at $d = 0.5$ as a function of matching correlation ρ: $\rho = 0.30 \rightarrow N = 55$ (adjusted); $\rho = 0.50 \rightarrow N = 39$ (adjusted); $\rho = 0.65 \rightarrow N = 26$ (adjusted, illustrative example used in this article); $\rho = 0.70 \rightarrow N = 23$ (adjusted). These values illustrate that the efficiency advantage over the RCT ($N = 151$ at 1:1) is maintained across a wide range of achievable matching correlations, reinforcing that the choice of $\rho = 0.65$ as an illustrative value does not drive the paper's core conclusions.

At $d = 0.5$, approximately 26 treated patients are required, representing a 5.8-fold reduction compared to RCT. A structured roadmap for integrating historical controls into clinical trials has been developed by the Drug Information Association Adaptive Design Scientific Working Group, providing practical guidance for regulatory submissions using this approach [17]. Propensity score matching, used here as one illustrative approach, is among the most widely used and transparent methods for reducing confounding in natural history comparator designs; alternative methods, including inverse probability of treatment weighting (IPTW), Mahalanobis distance matching, and regression-based covariate adjustment, are also acceptable and may be preferable depending on dataset characteristics [16].

TABLE 3 | Sample Size Requirements by Trial Design and Effect Size.

Effect Size (d)	RCT (1:1)/(2:1)	Natural History ($\rho = 0.65$)	Own-Control ($\rho = 0.90$)	Efficiency Gain
0.25	602/904	105	30	20×
0.50	151/226	26	8	19×
0.75	67/101	12	4	17×
1.00	38/57	7	2	19×

Note: Adjusted values with 20% inflation for practical considerations. Note that rarely would one use only 2 patients in a trial.

3.5 | Design 3: Patient-As-Own-Control

When each patient provides both pre- and post-treatment measurements, within-subject correlation for stable biological parameters is often high. For illustrative purposes, we use $\rho = 0.90$ as an example planning value. High within-patient stability for objective rare disease biomarkers has been empirically documented across multiple diseases: Within-subject variance of blood phenylalanine in PKU is substantially lower than between-subject variance across longitudinal serial measurements, consistent

17528062, 2026, 5, Downloaded from https://ascpt.onlinelibrary.wiley.com/doi/10.1111/cts.70551 by Readcube (Labtiva Inc.), Wiley Online Library on [28/10/2026]. 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

with intraclass correlations in the 0.70–0.90 range [18]; lyso-Gb3 in Fabry disease demonstrates intra-assay coefficients of variation below 10% with strong individual tracking over time [19]; and plasma chitotriosidase in Gaucher disease tracks individual treatment response with remarkable consistency over two decades of follow-up [20]. These data support the biological plausibility of within-patient correlations in this range for biochemical biomarkers in stable genetic diseases. However, ρ is critically disease- and endpoint-specific. For episodic outcomes (bleeding episodes, angioedema attacks, vaso-occlusive crises), event-count endpoints, or diseases with high spontaneous variability, ρ may be substantially lower and will depend on run-in duration. In such settings, sponsors must derive ρ empirically from natural history data rather than adopting a default assumption, and run-in periods should be long enough to establish a reliable baseline event rate. The mathematical framework is identical to matched designs, but ρ reflects the stability of individual baseline physiology rather than matching quality [21].

$$N_{Own} = 2(1-\rho)\left(Z_{\alpha/2}+Z_{\beta}\right)^2 / d^2$$

With $\rho = 0.90$:

$$N_{Own} = 1.57 / d^2(\text{base}) \rightarrow N_{Own} = 1.88 / d^2(\text{adjusted})$$

At $d = 0.5$, only 8 patients are required—a 19-fold reduction compared to RCT.

Patient-as-own-control designs are also constrained by endpoint type. Outcomes characterized by low event frequency, long latency to occurrence, or irreversibility—such as mortality, major organ failure, or catastrophic disease events—are poorly suited to within-patient comparison. For such endpoints, the pre-treatment observation window may not adequately capture the event rate needed to establish a reliable baseline, and the irreversibility of the outcome precludes the repeated measurement structure that underlies within-patient statistical efficiency. For example, survival in amyotrophic lateral sclerosis (ALS) or Duchenne muscular dystrophy cannot be treated as a pre-post paired measurement; vaso-occlusive crises in sickle cell disease occur with low and variable frequency, which challenges stable baseline estimation. In these settings, natural history comparator designs or adaptive trials with event-based endpoints are more appropriate alternatives.

The dramatic efficiency of within-subject designs stems from the simple mathematical fact that individual biological variation—the "noise" that obscures treatment signals in between-group comparisons—is eliminated when each patient serves as their own control. This is precisely why these designs are particularly powerful for genetically heterogeneous rare diseases: By comparing post-treatment to pre-treatment within the same genetic background, treatment effects emerge from what would otherwise be lost in inter-individual variation.

4 | When Supply Meets Demand: The Type II Error Problem

The collision between limited patient supply and inflexible trial design demands produces a predictable outcome: Systematic failure to detect real therapeutic effects. This is the Type II error problem—concluding that a treatment does not work when it actually does.

4.1 | The Mathematics of Underpowering

Consider a rare disease affecting 100 patients in the United States. A sponsor designs an RCT targeting an effect size of $d = 0.5$ (a clinically meaningful benefit). The calculation is straightforward:

- Required sample size (RCT): 151 patients
- Available patients: 100
- Realistic enrollment (accounting for eligibility, consent, access): ~50 patients

With 50 patients enrolled in an RCT design, statistical power drops below 50%. This means that even if the drug produces the targeted 0.5 standard deviation improvement, the trial has a greater than 50% probability of concluding "no effect." The trial is designed to fail before the first patient is enrolled if we stick with an alpha of less than 0.05.

Contrast this with a patient-as-own-control design using the same 50 patients. With $\rho = 0.90$, only 8 patients are needed for 80% power at $d = 0.5$. Fifty patients provide power exceeding 99% for this effect size, with adequate power to detect effects as small as $d = 0.2$—a four-fold improvement in sensitivity. Moreover, the efficiency of this design creates an opportunity unavailable in traditional RCTs: The ability to include a validation cohort. With only 8–10 patients required for the primary analysis under the illustrative $\rho = 0.90$ assumption, an alternative approach that maximizes the evidentiary value of the remaining patients is sequential analysis. In rare disease own-control trials, all available patients enroll concurrently—serial enrollment, waiting for each patient's results before enrolling the next, is operationally impractical. The sequential framework, therefore, applies to the analysis rather than to enrollment. After all patients have completed their pre- and post-treatment assessments, pre-post difference scores are ordered by a pre-specified analytical sequence, and a cumulative test statistic is computed by adding patients one at a time. Pre-specified stopping boundaries—such as an O'Brien-Fleming or Pocock boundary with an alpha-spending function—are defined in the statistical analysis plan before any data are unblinded [22]. If the boundary is crossed early in the analytical sequence, the remaining patients constitute a natural internal replication set; a second boundary crossing as additional patients are incorporated provides analytical confirmation within the same fully enrolled trial. If the boundary is never crossed after all patients are included, the trial is declared inconclusive. This property follows from the alpha-spending framework, which controls the overall Type I error rate regardless of how many analytical steps are performed, and each crossing is a genuine test of reproducibility [22]. A Bayesian sequential implementation, using the primary patients to generate a posterior distribution for the treatment effect and the remaining patients to update that posterior, provides an additional natural framework for quantifying the probability of a clinically meaningful effect [23]. Pre-specification of the

17528062, 2026, 5, Downloaded from https://ascpt.onlinelibrary.wiley.com/doi/10.1111/cts.70551 by Readcube (Labtiva Inc.), Wiley Online Library on [28/10/2026]. 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

analytical plan, including stopping boundaries, is essential for regulatory credibility—a luxury that underpowered RCTs, which are often simply declared negative when they fail to reach significance, cannot afford.

4.2 | One Chance to Succeed

The stakes in rare disease development extend beyond statistics. Unlike common disease programs, where failed trials can be redesigned and repeated, rare disease trials typically represent a single opportunity. The reasons are multiple:

- Limited sponsor resources: Many rare disease programs are conducted by small biotechnology companies, academic medical centers, or patient advocacy organizations operating with constrained budgets [24].
- Investor confidence: A negative Phase III or pivotal result, even if statistically underpowered, typically terminates future investment in the program.
- Patient community fatigue: Failed trials erode trust and reduce willingness to participate in future studies.

Given these realities, designing a rare disease trial with an inappropriate statistical model—one that cannot detect plausible effect sizes with the available patient population—represents both a missed opportunity and a failure of scientific stewardship. Patients participate in trials with the understanding that their contribution may help establish whether a therapy works; they deserve trial designs capable of answering that question.

4.3 | Maximizing Patient Inclusion

The scarcity of patients in rare diseases demands a fundamental rethinking of exclusion criteria. Traditional trials for common diseases routinely exclude patients based on age, comorbidities, prior treatments, and numerous other factors to create homogeneous study populations that simplify analysis. In rare diseases, such exclusions may eliminate the majority of available patients.

We advocate for minimal exclusion criteria in rare disease trials, accepting the heterogeneity that this introduces. The patient-as-own-control design is particularly suited to this approach precisely because heterogeneity between subjects does not contaminate the within-subject treatment effect estimate. A patient with atypical disease presentation or comorbidities can still serve as their own valid control, provided their baseline measurements reflect stable disease or a quantifiable disease trajectory.

Furthermore, pediatric inclusion must become the default rather than the exception. With 75% of rare diseases manifesting in childhood, trials that exclude children test therapies in the wrong population at the wrong point in disease progression. The efficiency gains from within-subject designs can facilitate the smaller sample sizes and reduced placebo exposure that make pediatric trials ethically feasible.

4.4 | Post-Study Analysis Methods: Matching Design to Analysis

The choice of trial design directly influences the appropriate analytical framework for demonstrating treatment efficacy. For natural history comparator designs, propensity score matching is among the most widely used and transparent methods for reducing confounding between treated and historical control populations; alternative approaches, including inverse probability of treatment weighting (IPTW), Mahalanobis distance matching, and regression-based covariate adjustment, are also acceptable depending on dataset characteristics [16]. This approach creates balanced comparison groups by matching on observed baseline covariates—age at diagnosis, disease severity, genetic subtype, and baseline biomarker levels—thereby approximating the balance achieved through randomization [16]. The strength of propensity score methods lies in their transparency and reproducibility; the weakness is their inability to account for unmeasured confounders, making the quality of the natural history dataset paramount. Sensitivity analyses using multiple matching algorithms (nearest neighbor, caliper matching, inverse probability weighting) can assess the robustness of findings. When longitudinal natural history data are available, linear mixed models (LMMs) can characterize individual disease trajectories in the control population, enabling comparison of treated patients' outcomes against their predicted untreated course—a particularly powerful approach when disease progression rates vary substantially among patients [25]. However, the accuracy of predicted trajectories depends critically on phenotypic heterogeneity: When within-disease variability is high, predicted individual trajectories carry wide confidence intervals that reduce the precision of the treatment effect estimate. The random effects structure must be validated against the natural history data before predictive use, and sponsors should prespecify sensitivity analyses using alternative trajectory specifications—including non-linear and piecewise linear models for diseases with identifiable inflection points—to establish robustness. When the natural history dataset is small, variance components for random slopes may be poorly estimated regardless of model fit.

For patient-as-own-control designs, the paired *t*-test or its nonparametric equivalent (Wilcoxon signed-rank test) provides the simplest analytical framework, comparing each patient's posttreatment outcome to their own baseline. The strength of this approach is its simplicity and the elimination of between-subject variability; the weakness is susceptibility to time-dependent confounders such as disease progression, placebo effects, or regression to the mean. For designs with multiple assessment timepoints, linear mixed models offer substantial advantages: They accommodate unbalanced data when patients miss visits, model individual response trajectories through random slopes and intercepts, and properly account for the correlation structure of repeated measurements within subjects [26]. Mixed-effects models for repeated measures (MMRM) represent a specific LMM implementation widely accepted by regulatory agencies, allowing characterization of treatment effect trajectories while providing valid inference under the missing-at-random (MAR) assumption as the primary analytical framework [27]. However, regulators require pre-specified sensitivity

17528062, 2026, 5, Downloaded from https://ascpt.onlinelibrary.wiley.com/doi/10.1111/cts.70551 by Readcube (Labtiva Inc.), Wiley Online Library on [28/10/2026]. 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

analyses under missing-not-at-random (MNAR) assumptions—including tipping point analyses and pattern mixture models—to establish the robustness of findings to different missingness mechanisms. In open-label run-in designs, where dropout may be informatively related to treatment response (patients who improve dramatically may feel less need to continue, while patients who deteriorate rapidly may be unable to continue), the case for MNAR sensitivity analyses is particularly strong and should be addressed proactively in the statistical analysis plan.

For ultra-rare diseases where even these designs yield very small samples, Bayesian analytical frameworks offer particular advantages: They can formally incorporate prior information from preclinical studies, mechanistic understanding, or related diseases to augment the limited trial data, and they provide probability statements about treatment effects that are often more clinically interpretable than traditional *p*-values [23]. Regardless of design, pre-specification of the analytical plan—including primary analysis, sensitivity analyses, and handling of missing data—remains essential for regulatory credibility. Additionally, pre-specification of what constitutes clinically meaningful improvement or stability adds rigor. In certain designs, these can be designated on a per-patient basis.

5 | Addressing Bias and the Role of Objective Endpoints

A legitimate concern with alternative trial designs is their potential vulnerability to bias. Without concurrent randomization, patient-as-own-control and natural history comparator designs must address regression to the mean, spontaneous disease fluctuation, selection bias, and temporal confounding. However, the magnitude of these concerns depends critically on endpoint selection—and for most rare genetic diseases characterized by objective, measurable outcomes and relatively stable or predictable disease trajectories, whether or not the etiology is genetic, these biases are substantially mitigated. Approximately 72% of rare diseases have genetic origins and therefore tend to feature objective biochemical or functional biomarkers, but the design framework presented here applies equally to non-genetic rare diseases when endpoints meet the same objectivity criteria.

5.1 | The Placebo Effect Is Endpoint-Dependent

A landmark systematic review by Hróbjartsson and Gøtzsche examined 130 trials comparing placebo to no treatment and found that placebo effects are concentrated almost exclusively in subjective outcomes [28]. For continuous subjective endpoints (pain, fatigue, quality of life), the standardized mean difference was −0.36 (95% CI: −0.47 to −0.25). For continuous objective endpoints, however, the effect was −0.12 (95% CI: −0.27 to 0.03)—not statistically significant. For binary outcomes regardless of subjectivity, placebo showed no effect (RR 0.95, 95% CI: 0.88–1.02).

This finding has profound implications for rare disease research. Approximately 72% of rare diseases have genetic origins, typically involving enzyme deficiencies, substrate accumulation, or structural protein abnormalities that produce objectively measurable pathology. A patient with Gaucher disease cannot "believe" their glucocerebrosidase activity into normalcy; globotriaosylceramide deposition in Fabry disease is quantified by tissue biopsy, not patient perception; blood phenylalanine in PKU is a biochemical measurement uninfluenced by expectation. In addition, the use of new technologies like video scoring and observation may improve the objectivity of the more subjective observations. Where objective measurements are not possible, randomized placebo run-in periods or other designs can be used in self-controlled studies.

5.2 | Regulatory Acceptance of Objective Biomarker Endpoints

The FDA has explicitly recognized this principle in approving numerous rare disease therapies based on biomarker endpoints: Agalsidase beta (Fabry disease) using GL3 clearance on renal biopsy [30]; sapropterin (PKU) using blood phenylalanine levels [31]; sodium phenylbutyrate (urea cycle disorders) using plasma ammonia control [32]; and alglucosidase alfa (Pompe disease) using survival and ventilator-free survival [29]. Regulatory guidance for qualifying biomarkers as primary endpoints under the accelerated approval pathway further supports this approach [33]. Of 45 FDA approvals using external controls between 2000 to 2019, 44% employed retrospective natural history data, demonstrating regulatory acceptance when disease trajectory is well-characterized, and endpoints are objective [34].

5.3 | When Bias Matters—And When It Does Not

Bias remains a legitimate concern when endpoints are subjective, the disease course is unpredictable with spontaneous remissions, or treatment effects are subtle. For such conditions—chronic pain, psychiatric disorders, relapsing–remitting diseases—RCTs retain their methodological value. However, most rare genetic diseases do not fit this profile. They feature unidirectional progression without spontaneous improvement, objectively measurable pathology, and well-characterized natural history. For these diseases, the theoretical vulnerabilities of alternative designs are mitigated by endpoint selection, while the practical impossibility of adequately powering RCTs makes alternative designs not merely acceptable but necessary.

Propensity score matching can further address selection bias in natural history comparator designs, reducing systematic differences between treated and historical control groups, as discussed in the Design 2 section. Combined with blinded independent central review for endpoint adjudication and prespecified sensitivity analyses, these methods provide scientific rigor comparable to randomized designs at a fraction of the sample size requirement.

6 | Case Studies: High-Profile RCT Failures and Alternative Designs

The following late-stage clinical trial failures illustrate how inappropriate trial design challenges contributed to negative outcomes in rare disease programs. These cases represent a

17528062, 2026, 5, Downloaded from https://ascpt.onlinelibrary.wiley.com/doi/10.1111/cts.70551 by Readcube (Labtiva Inc.), Wiley Online Library on [28/10/2026]. 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

TABLE 4 | Late-Stage RCT Failures in Rare Diseases and Potential Alternative Designs.

Drug (Sponsor)	Disease	Failure Mode	Alternative Design
Fordadistrogene movaparvovec (Pfizer) [35]	DMD	Failed NSAA primary endpoint	Run-in design assessing individual response
Pamrevlumab (FibroGen) [36]	DMD	Non-significant motor improvement	Crossover design enhancing power
AMX0035 (Amylyx) [37]	ALS	No significant ALSFRS-R change	Own-control reducing variability
Memantine (multiple) [38]	ALS	Failed to show efficacy	Within-patient design for progression
Talampanel (multiple) [39]	ALS	Failed in advanced ALS	Natural history controls for survival
Tominersen (Roche) [40]	Huntington's	Patient stratification and dose-exposure problem: Higher drug exposure harmed older patients with greater disease burden while younger patients with lower disease burden showed potential benefit in post hoc analyses; adequate enrollment ($N = 791$) but inappropriate single-dose frequency for all comers	Adaptive design with pre-specified subgroup stratification by age, disease burden (CAP score), and dosing frequency; would enable dose-finding within defined patient subgroups and prospective testing of the younger/lower-burden hypothesis now being tested in GENERATION HD2
Catena (Santhera) [41]	Friedreich's Ataxia	Failed primary and secondary endpoints	Own-control design
Actimmune (Horizon) [42]	Friedreich's Ataxia	Failed to improve clinical scores	Historical control comparison
AT-GAA (Amicus) [43]	Pompe Disease	Did not meet endpoints	Run-in baseline comparison
Losmapimod (Fulcrum) [44]	FSHD	Failed primary functional endpoint (Reachable Workspace)	Biomarker-driven adaptive design
Labafenogene marselecobac (Synlogic) [45]	PKU	Failed efficacy thresholds	Within-patient metabolic comparison
Rivipansel (Pfizer) [46]	Sickle Cell	No benefit in VOC duration	Natural history control comparison
Pitolisant (Harmony Biosciences/Jazz)—INTUNE trial [47]	Idiopathic Hypersomnia	Failed primary endpoint (Epworth Sleepiness Scale) in Phase 3 INTUNE trial; subjective patient-reported sleepiness measure vulnerable to placebo effects in open-label extension; inadequate differentiation from placebo on primary endpoint	Boundary case illustrating limits of own-control applicability: Subjective sleepiness endpoints create substantial placebo vulnerability, making standard own-control designs less appropriate. An adaptive crossover design with pre-specified objective polysomnographic endpoints (sleep latency, total sleep time) alongside patient-reported outcomes might have provided more tractable objective signal while preserving the within-patient efficiency advantage. Alternatively, a natural history comparator using actigraphy-based sleep metrics could separate true drug effect from subjective report bias.

Note: See Table 2 for statistical term definitions.
Abbreviations: ALS = amyotrophic lateral sclerosis; ALSFRS-*R* = ALS Functional Rating Scale-Revised; CAP = CAG-Age Product; DMD = Duchenne muscular dystrophy; FSHD = facioscapulohumeral muscular dystrophy; NSAA = North Star Ambulatory Assessment; PKU = phenylketonuria; VOC = vaso-occlusive crisis. RCT = randomized controlled trial.

17528062, 2026, 5, Downloaded from https://ascpt.onlinelibrary.wiley.com/doi/10.1111/cts.70551 by Readcube (Labtiva Inc.), Wiley Online Library on [28/10/2026]. 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

heterogeneous mix of failure modes—statistical power limitations, patient population stratification issues, endpoint selection challenges, and true biological signal absence—and the applicability of alternative designs varies accordingly. For those trials where enrolled N fell substantially below what was needed to detect plausible effect sizes, the power argument is strongest; for others, such as tominersen, the primary challenge was patient stratification rather than insufficient enrollment (Table 4).

These thirteen failures span neuromuscular diseases (DMD, ALS, facioscapulohumeral muscular dystrophy (FSHD)), neurological conditions (Huntington's, Friedreich's Ataxia), metabolic disorders (Pompe, PKU), and hematological disease (Sickle Cell)—representing billions of dollars in lost investment, years of delayed therapy access, and continued suffering for patients. While we cannot know with certainty whether alternative designs would have yielded different outcomes, the mathematics strongly suggest that statistical power limitations contributed to at least some of these negative results. Available evidence from the general clinical development literature and rare disease trial analyses consistently identifies recruitment and retention failures, inadequate statistical power, and operational challenges as leading contributors to trial failure—with true lack of biological activity representing a smaller proportion [5–8]. This distinction is critical: Many potentially effective therapies are abandoned not because they do not work, but because the trial design selected could not detect whether they worked.

7 | Alternative Trial Design Successes: FDA and EMA Approvals

The following tables document successful drug approvals achieved using alternative trial designs—natural history controls and patient-as-own-control approaches—demonstrating that regulatory agencies accept and approve drugs based on these methodologically rigorous alternatives to traditional RCTs (Tables 5 and 6).

TABLE 5 | Natural History Control Design Successes.

Drug (Sponsor)	Disease	Design	Year	N	Center
Myozyme (Genzyme) [29]	Pompe disease	Single-arm open-label vs. natural history controls	2006	18	CBER
Zolgensma (Novartis/AveXis) [48]	Spinal muscular atrophy	Single-arm open-label vs. natural history data	2019	36	CBER
Nulibry (BridgeBio/Origin) [49]	Molybdenum cofactor deficiency Type A	Single-arm vs. genotype-matched natural history	2021	13 treated vs. 18 controls	CDER
Skysona (Bluebird Bio) [50]	Cerebral adrenoleukodystrophy	Single-arm gene therapy vs. natural history	2022	32	CBER
Zokinvy (Eiger) [51]	Hutchinson-Gilford progeria syndrome	Single-arm vs. historical untreated survival data	2020	62	CDER
Qalsody (Biogen) [52]	ALS with SOD1 mutation	Single-arm with biomarker + natural history	2023	108	CDER
Sohonos (palovarotene) [53]	Fibrodysplasia ossificans progressiva	Single-arm vs. natural history bone formation	2023	107	CDER

TABLE 6 | Patient-as-Own-Control Design Successes (Baseline Comparison).

Drug (Sponsor)	Disease	Design	Year	N	Center
Veopoz (Regeneron) [54]	CHAPLE disease (CD55 deficiency)	Single-arm, baseline comparison	2023	10	CDER
Imcivree (Rhythm) [55, 56]	POMC, PCSK1, or LEPR deficiency obesity	Single-arm, baseline weight comparison	2020	21	CDER
Dojolvi (Ultragenyx) [57]	Long-chain fatty acid oxidation disorders	Single-arm, pre/post comparison	2020	29	CDER
Kebilidi (PTC Therapeutics) [58]	AADC deficiency	Single-arm gene therapy, milestone achievement from baseline	2024	35	CBER
Casgevy (Vertex/CRISPR) [59]	Sickle cell disease	Single-arm gene therapy, VOC-free vs. baseline	2023	44	CBER
Gamifant (Sobi) [60]	Primary HLH	Single-arm, response rate from baseline	2018	27	CDER

17528062, 2026, 5, Downloaded from https://ascpt.onlinelibrary.wiley.com/doi/10.1111/cts.70551 by Readcube (Labtiva Inc.), Wiley Online Library on [28/10/2026]. 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

7.1 | Key Observations

These 13 approvals demonstrate several critical points. Both CDER and CBER accept alternative designs, with 8 approvals through CDER and 5 through CBER. Ultra-rare diseases predominate, with most diseases having fewer than 1000 patients worldwide. Sample sizes are small but adequate, ranging from 10 (Veopoz) to 108 (Qalsody), with a median of approximately 32. Natural history controls were accepted when the disease course is well-characterized, no spontaneous improvement occurs, and a strong biologic rationale exists. Baseline comparisons were accepted when endpoints are objective, the disease is stable without treatment, and changes are measurable and clinically meaningful. Gene therapies frequently use alternative designs, as exemplified by Zolgensma, Skysona, Casgevy, and Kebilidi, all of which used single-arm designs.

8 | Conclusions and Recommendations

The mathematics are unambiguous: Traditional RCT designs are fundamentally mismatched to rare disease research. For >80% of rare diseases, the global patient population falls below the sample size required for conventional trials to detect clinically meaningful effects. Continuing to apply these designs ensures that most rare disease trials will be underpowered, many will fail, and effective therapies will never reach patients.

Alternative designs—particularly patient-as-own-control and natural history comparator models—offer mathematically rigorous solutions that reduce sample size requirements by 5- to 20-fold while preserving statistical validity. These designs are particularly well-suited to rare diseases because they: control for genetic heterogeneity by using each patient as their own baseline; facilitate pediatric enrollment by reducing total patient requirements and minimizing placebo exposure; enable broader inclusion criteria by tolerating inter-subject heterogeneity; and maximize information extraction from each irreplaceable patient participant.

We call upon regulatory agencies to formalize acceptance of these designs through explicit guidance documents, with clear statistical frameworks specifying acceptable correlation assumptions and analytical approaches. We urge sponsors to consider trial design as the primary determinant of success or failure, selecting designs based on mathematical feasibility rather than conventional precedent. And we encourage patient advocacy organizations to demand transparency regarding the statistical power of proposed trials—patients deserve to know whether the trials they join are capable of detecting treatment effects. The field must prioritize trial designs that maximize the probability of detecting true treatment effects over those that merely satisfy conventional regulatory expectations.

The 90% of rare diseases without approved treatments represent not merely a therapeutic gap but a methodological failure. We have the statistical tools to close this gap. What remains is the willingness to use them.

Acknowledgments

The authors thank the rare disease patient communities whose experiences informed this analysis.

Use of Artificial Intelligence: In accordance with Clinical and Translational Science guidelines, the authors acknowledge the use of AI-assisted tools (Claude, Anthropic) in the preparation of this manuscript. AI assistance was limited to editorial support functions including reference formatting, grammatical review, consistency checking across sections, and table formatting. All scientific concepts, intellectual arguments, analytical frameworks, data presentations, statistical derivations, clinical interpretations, and policy recommendations presented in this manuscript originate solely from the authors. The authors take full responsibility for the accuracy, integrity, and content of all work submitted.

Funding

The authors have nothing to report.

Conflicts of Interest

The authors declare no conflicts of interest.

References

1. Medical Research Council, "Streptomycin Treatment of Pulmonary Tuberculosis: A Medical Research Council Investigation," *BMJ Clinical Research Education* 2, no. 4582 (1948): 769–782.

2. J. S. Bailly, B. Franklin, A. Lavoisier, et al., "Rapport des commissaires chargés par le roi de l'examen du magnétisme animal. 1784," in (Paris: Imprimerie Royale).

3. J. Shourick, F. Ruiz, S. Aymé, et al., "Estimating the Prevalence of Rare Diseases: A Critical Review of Literature Methods," *Orphanet Journal of Rare Diseases* 16 (2021): 139.

4. S. Nguengang Wakap, D. M. Lambert, A. Olry, et al., "Estimating Cumulative Point Prevalence of Rare Diseases: Analysis of the Orphanet Database," *European Journal of Human Genetics* 28 (2020): 165–173.

5. D. B. Fogel, "Factors Associated With Clinical Trials That Fail and Opportunities for Improving the Likelihood of Success: A Review," *Contemporary Clinical Trials Communications* 11 (2018): 156–164.

6. R. B. Gul and P. A. Ali, "Clinical Trials: The Challenge of Recruitment and Retention of Participants," *Journal of Clinical Nursing* 19 (2010): 227–233.

7. S. A. Bell and C. Tudur Smith, "A Comparison of Interventional Clinical Trials in Rare Versus Non-Rare Diseases: An Analysis of ClinicalTrials.Gov," *Orphanet Journal of Rare Diseases* 9 (2014): 170.

8. S. W. Hee, A. Willis, C. Tudur Smith, et al., "Does the Low Prevalence Affect the Sample Size of Interventional Clinical Trials of Rare Diseases? An Analysis of Data From the Aggregate Analysis of ClinicalTrials.Gov," *Orphanet Journal of Rare Diseases* 12 (2017): 44.

9. C. R. Ferreira, "The Burden of Rare Diseases," *American Journal of Medical Genetics. Part A* 179 (2019): 885–892.

10. D. Taruscio, M. Salvatore, P. Torreri, et al., "Undiagnosed Diseases: Italy-US Collaboration to Promote Research and Patient Care," *International Journal of Environmental Research and Public Health* 20 (2023): 5570.

11. Orphanet, *Prevalence of rare diseases: Bibliographic data*. Orphanet Report Series, 2024 accessed September 2024, https://www.orpha.net.

17528062, 2026, 5, Downloaded from https://ascpt.onlinelibrary.wiley.com/doi/10.1111/cts.70551 by Readcube (Labtiva Inc.), Wiley Online Library on [28/10/2026]. 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

12. A. Y. L. Chan, V. K. Y. Chan, S. Olsson, et al., "Access and Unmet Needs of Orphan Drugs in 194 Countries and 6 Areas: A Global Policy Review With Content Analysis," *Value in Health* 23 (2020): 1580–1591.

13. J. A. Cook, S. A. Julious, W. Sones, et al., "DELTA2 Guidance on Choosing the Target Difference and Undertaking and Reporting the Sample Size Calculation for a Randomised Controlled Trial," *Health Technology Assessment* 22 (2018): 1–56.

14. J. Cohen, *Statistical Power Analysis for the Behavioral Sciences, 2nd ed.* (Hillsdale, NJ: Lawrence Erlbaum Associates, 1988).

15. S. J. Pocock, "The Combination of Randomized and Historical Controls in Clinical Trials," *Journal of Chronic Diseases* 29, no. 3 (1976): 175–188.

16. P. C. Austin, "An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies," *Multivariate Behavioral Research* 46, no. 3 (2011): 399–424.

17. M. Ghadessi, R. Tang, J. Zhou, et al., "A Roadmap to Using Historical Controls in Clinical Trials—By Drug Information Association Adaptive Design Scientific Working Group (DIA-ADSWG)," *Orphanet Journal of Rare Diseases* 15 (2020): 69.

18. B. K. Burton, H. Bausell, R. Katz, H. LaDuca, and C. Sullivan, "Sapropterin Therapy Increases Stability of Blood Phenylalanine Levels in Patients With BH4-Responsive Phenylketonuria (PKU)," *Molecular Genetics and Metabolism* 101, no. 2–3 (2010): 110–114.

19. U. Ramaswami, M. L. West, K. Tylee, et al., "The Use and Performance of Lyso-Gb3 for the Diagnosis and Monitoring of Fabry Disease: A Systematic Literature Review," *Molecular Genetics and Metabolism* 145, no. 2 (2025): 109110.

20. P. Szymańska-Rożek, B. Czartoryska, G. Kleinotiene, P. Lipiński, A. Tylki-Szymańska, and A. Ługowska, "A 20-Year Longitudinal Study of Plasma Chitotriosidase Activity in Treated Gaucher Disease Type 1 and 3 Patients—A Qualitative and Quantitative Approach," *Biomolecules* 13, no. 3 (2023): 436.

21. S. Gupta, M. E. Faughnan, L. Engel, et al., "Recommendations for Clinical Trial Design in Patients as Their Own Control in Rare Diseases," *Orphanet Journal of Rare Diseases* 16 (2021): 308.

22. C. Jennison and B. W. Turnbull, *Group Sequential Methods With Applications to Clinical Trials.* (Boca Raton: Chapman & Hall/CRC, 1999).

23. L. V. Hampson, J. Whitehead, D. Eleftheriou, and P. Brogan, "Bayesian Methods for the Design and Interpretation of Clinical Trials in Very Rare Diseases," *Statistics in Medicine* 33, no. 24 (2014): 4186–4201.

24. K. Jayasundara, A. Hollis, M. Krahn, M. Mamdani, J. S. Hoch, and P. Grootendorst, "Estimating the Clinical Cost of Drug Development for Orphan Versus Non-Orphan Drugs," *Orphanet Journal of Rare Diseases* 14 (2019): 12.

25. N. M. Laird and J. H. Ware, "Random-Effects Models for Longitudinal Data," *Biometrics* 38, no. 4 (1982): 963–974.

26. G. Verbeke and G. Molenberghs, *Linear Mixed Models for Longitudinal Data,* (New York: Springer, 2000).

27. C. H. Mallinckrodt, W. S. Clark, R. J. Carroll, et al., "Recommendations for the Primary Analysis of Continuous Endpoints in Longitudinal Clinical Trials," *Drug Information Journal* 42, no. 4 (2008): 303–319.

28. A. Hróbjartsson and P. C. Gøtzsche, "Is the Placebo Powerless? An Analysis of Clinical Trials Comparing Placebo With no Treatment," *New England Journal of Medicine* 344, no. 21 (2001): 1594–1602.

29. P. S. Kishnani, D. Corzo, M. Nicolino, et al., "Recombinant Human Acid α-Glucosidase: Major Clinical Benefits in Infantile-Onset Pompe Disease," *Neurology* 68, no. 2 (2007): 99–109.

30. C. M. Eng, N. Guffon, W. R. Wilcox, et al., "Safety and efficacy of recombinant human alpha-galactosidase A replacement therapy in Fabry's disease," *New England Journal of Medicine* 345, no. 1 (2001): 9–16.

31. H. L. Levy, A. Milanowski, A. Chakrapani, et al., "Efficacy of sapropterin dihydrochloride (tetrahydrobiopterin, 6R-BH4) for reduction of phenylalanine concentration in patients with phenylketonuria: a phase III randomised placebo-controlled study," *Lancet* 370, no. 9586 (2007): 504–510.

32. G. M. Enns, S. A. Berry, G. M. Berry, W. J. Rhead, S. W. Brusilow, A. Hamosh et al., "Survival after treatment with phenylacetate and benzoate for urea-cycle disorders," *New England Journal of Medicine* 356, no. 22 (2007): 2282–2292.

33. U.S. Food and Drug Administration, *Accelerated Approval of Drugs and Biologics for Serious Conditions: Guidance for Industry.* (Silver Spring, MD: FDA, 2023), available at https://www.fda.gov/regulatory-information/search-fda-guidance-documents/accelerated-approval-drugs-and-biologics-serious-conditions.

34. J. P. Jarow, L. LaVange, and J. Woodcock, "The Use of External Controls in FDA Regulatory Decision Making," *Therapeutic Innovation & Regulatory Science* 55 (2021): 1002–1009.

35. Pfizer Inc, Pfizer provides update on Phase 3 gene therapy study for ambulatory boys with Duchenne muscular dystrophy. 2023.

36. FibroGen Inc, "FibroGen reports topline results from Phase 3 LELANTOS-2 trial of pamrevlumab in Duchenne muscular dystrophy." 2023. Press release

37. Amylyx Pharmaceuticals, Amylyx Pharmaceuticals announces topline results from PHOENIX trial evaluating AMX0035 in ALS. 2024.

38. M. de Carvalho, S. Pinto, J. Costa, T. Evangelista, B. Ohana, and A. Pinto, "A Randomized, Placebo-Controlled Trial of Memantine for Functional Disability in Amyotrophic Lateral Sclerosis," *Amyotrophic Lateral Sclerosis* 11, no. 5 (2010): 456–460.

39. R. M. Pascuzzi, J. Shefner, A. S. Chappell, et al., "A Phase II Trial of Talampanel in Subjects With Amyotrophic Lateral Sclerosis," *Amyotrophic Lateral Sclerosis* 11, no. 3 (2010): 266–271.

40. P. McColgan, S. J. Tabrizi, R. S. Doody, and GENERATION HD1 Investigators, "Tominersen in Adults With Manifest Huntington's Disease," *New England Journal of Medicine* 389, no. 23 (2023): 2203–2205.

41. D. R. Lynch, S. L. Perlman, and T. Meier, "A Phase 3, Double-Blind, Placebo-Controlled Trial of Idebenone in Friedreich Ataxia," *Archives of Neurology* 67, no. 8 (2010): 941–947.

42. L. A. Seyer, K. Galetta, J. Wilson, et al., "Analysis of the Safety and Efficacy of Interferon Gamma-1b in Friedreich Ataxia," *Journal of Neuroimmunology* 284 (2015): 1–4.

43. Amicus Therapeutics, Amicus Therapeutics reports third quarter 2022 financial results and provides business update. 2022.

44. Fulcrum Therapeutics, "Fulcrum Therapeutics reports topline results from Phase 3 ReDUX4 study of losmapimod in facioscapulohumeral muscular dystrophy." 2024. Press release

45. Synlogic Inc, "Synlogic discontinues development of labafenogene marselecobac for PKU following Phase 3 failure." 2023. Press release

46. Pfizer Inc, "Pfizer Announces Phase 3 Top-Line Results for Rivipansel in Patients with Sickle Cell Disease Experiencing Vaso-Occlusive Crisis." 2019. Press release

47. Jazz Pharmaceuticals, Jazz Pharmaceuticals announces results from INTUNE Phase 3 study evaluating pitolisant for idiopathic hypersomnia. Press release. 2023.

48. J. R. Mendell, S. Al-Zaidy, R. Shell, et al., "Single-Dose Gene-Replacement Therapy for Spinal Muscular Atrophy," *New England Journal of Medicine* 377, no. 18 (2017): 1713–1722.

17528062, 2026, 5, Downloaded from https://ascpt.onlinelibrary.wiley.com/doi/10.1111/cts.70551 by Readcube (Labtiva Inc.), Wiley Online Library on [28/10/2026]. 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

49. S. Farrell, T. Hartley, A. Bhatt, et al., "Regulatory News: Nulibry (Fosdenopterin) Approved to Reduce the Risk of Mortality in Patients With Molybdenum Cofactor Deficiency Type A: FDA Approval Summary," *Journal of Inherited Metabolic Disease* 44, no. 5 (2021): 1085–1087.

50. bluebird bio Inc, bluebird bio receives FDA accelerated approval for SKYSONA gene therapy for early, active cerebral adrenoleukodystrophy. Press release, 2022.

51. L. B. Gordon, H. Shappell, J. Massaro, et al., "Association of Lonafarnib Treatment vs no Treatment With Mortality Rate in Patients With Hutchinson-Gilford Progeria Syndrome," *Journal of the American Medical Association* 319, no. 16 (2018): 1687–1695.

52. T. M. Miller, M. E. Cudkowicz, A. Genge, et al., "Trial of Antisense Oligonucleotide Tofersen for SOD1 ALS," *New England Journal of Medicine* 387, no. 12 (2022): 1099–1110.

53. US Food and Drug Administration, *Summary Basis of Regulatory Action: Sohonos (Palovarotene)* (Silver Spring, MD: FDA, 2023).

54. US Food and Drug Administration, *Drug Approval Package: Veopoz (pozelimab-bbfg)*. (Silver Spring, MD, 2023).

55. K. Clément, H. Biebermann, I. S. Farooqi, et al., "MC4R Agonism Promotes Durable Weight Loss in Patients With Leptin Receptor Deficiency," *Nature Medicine* 24, no. 5 (2018): 551–555.

56. K. Clément, E. L. T. van den Akker, J. Argente, et al., "Efficacy and safety of setmelanotide, an MC4R agonist, in individuals with severe obesity due to LEPR or POMC deficiency," *Lancet Diabetes & Endocrinology* 8, no. 12 (2020): 960–970.

57. J. Vockley, B. Burton, G. T. Berry, et al., "Triheptanoin Treatment in Patients With Pediatric Cardiomyopathy Associated With Long Chain-Fatty Acid Oxidation Disorders," *Molecular Genetics and Metabolism* 119, no. 3 (2016): 223–231.

58. C. H. Tai, N. C. Lee, Y. H. Chien, et al., "Long-Term Efficacy and Safety of Eladocagene Exuparvovec in Patients With AADC Deficiency," *Molecular Therapy* 30, no. 2 (2022): 509–518.

59. H. Frangoul, D. Altshuler, M. D. Cappellini, et al., "CRISPR-Cas9 Gene Editing for Sickle Cell Disease and β-Thalassemia," *New England Journal of Medicine* 384, no. 3 (2021): 252–260.

60. F. Locatelli, M. B. Jordan, C. Allen, et al., "Emapalumab in Children With Primary Hemophagocytic Lymphohistiocytosis," *New England Journal of Medicine* 382, no. 19 (2020): 1811–1822.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Table S1:** Sensitivity of Required Sample Size to Matching Correlation (ρ) for Natural History Comparator Design.

17528062, 2026, 5, Downloaded from https://ascpt.onlinelibrary.wiley.com/doi/10.1111/cts.70551 by Readcube (Labtiva Inc.), Wiley Online Library on [28/10/2026]. 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