Check for updates

Clinical and Translational Science

REVIEW OPEN ACCESS

The Placebo Effect in Rare Disease Clinical Trials: Measurement, Impact, and Statistical Approaches for Patient-as-Own-Control Designs

Marshall L. Summar[1,2] | Janet Woodcock[3]

[1]George Washington University, Washington, DC, USA | [2]Uncommon Cures, Chevy Chase, Maryland, USA | [3]Center for Drug Evaluation and Research, U.S. Food and Drug Administration, Brookeville, Maryland, USA

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

Received: 19 February 2026 | **Revised:** 23 March 2026 | **Accepted:** 26 March 2026

ABSTRACT

A frequently cited concern regarding patient-as-own-control trial designs in rare disease is the potential for placebo and related effects to inflate apparent treatment efficacy. Whether this concern is disqualifying or manageable has not been systematically evaluated. We reviewed meta-analyses quantifying placebo effect magnitude by endpoint type, reporter modality, and trial duration and evaluated statistical methods available for post-trial placebo adjustment in own-control designs. Placebo effects depend heavily on endpoint type. For objective endpoints (enzyme activity, serum biomarkers, imaging volumetrics)—which constitute the majority of primary endpoints in approved rare disease therapies—placebo effects are consistently small and in most meta-analyses statistically indistinguishable from zero (standardized mean difference [SMD] < 0.10). For subjective endpoints (patient-reported pain, caregiver-rated function), effects are larger (SMD 0.20–0.50) but well-characterized and correctable. Placebo effects peak early and decay over weeks, providing a temporal signature distinguishable from sustained pharmacological effects. Multiple complementary analytical methods—including temporal trajectory modeling, objective–subjective concordance analysis, Bayesian informative priors, extended run-in observation designs, and blinded outcome assessment—are available to quantify and manage placebo contributions. Importantly, the randomized controlled trial's structural advantage in canceling placebo is itself degraded in small samples, where asymmetric allocation of placebo responders can distort between-arm comparisons. The placebo objection to own-control designs is manageable rather than disqualifying. For objective endpoints, correction is minimal. For subjective endpoints, a rich analytical toolkit supports credible decomposition of drug and placebo components. These findings support the broader adoption of own-control designs in rare disease clinical trials.

1 | Introduction

Traditional double-blind placebo-controlled randomized controlled trials (DBPCRCTs) demand patient numbers that exceed the total available population for more than 95% of rare diseases [1]. Alternative 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, as we have detailed in the companion article to this review [1] Specifically, a patient-as-own-control design with within-subject correlation $\rho = 0.90$ reduces total required patients approximately 19-fold relative to a parallel-group RCT at 80% power ($\alpha = 0.05$, two-tailed) for a continuous endpoint with equal variance; these figures are provided across a range of effect sizes in table 3 of Summar and Woodcock [1]. This challenge is compounded by evidence that target differences in RCTs are often inadequately justified [2], and that the factors most associated with clinical trial failure—enrollment difficulties,

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

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

underpowering, and endpoint selection—are all amplified in rare disease settings [3].

A common objection to open-label and patient-as-own-control designs is that they are susceptible to placebo effects and related biases that could inflate apparent treatment efficacy. This concern is legitimate: own-control designs lack the concurrent control arm that structurally cancels placebo in the RCT, they assume baseline stability that may not hold in all diseases, and they require additional analytical steps to isolate the true drug effect. The concern is raised frequently enough that it warrants systematic examination. In this article, we address the placebo effect directly: what it is, how large it is, where it matters most, and—critically—how it can be quantified and managed in the post-trial analysis of own-control studies.

An important distinction frames our entire analysis: the difference between objective and subjective outcome measures. Objective endpoints—enzyme activity levels, serum biomarker concentrations, imaging volumetrics, hematologic parameters, survival—are determined by laboratory instruments, imaging technology, or unambiguous clinical events. A patient's expectations cannot meaningfully change their acid alpha-glucosidase activity or their low-density lipoprotein (LDL)-cholesterol level. Subjective endpoints—patient-reported pain, fatigue, quality of life, caregiver-rated functional scales—depend on human perception and judgment, which are susceptible to expectation and context. As we will show, the placebo literature consistently demonstrates that placebo effects are a relevant consideration primarily for subjective endpoints, while objective endpoints are minimally affected regardless of trial design.

A central point frames this discussion: placebo effects occur in every clinical trial design, including DBPCRCTs. In a blinded RCT, patients in both arms experience genuine placebo responses—improvements driven by expectation, conditioning, and the therapeutic relationship [4, 5]. The RCT handles this elegantly: because both arms experience the same placebo effect, it cancels when the between-group difference is computed. The result directly reveals d_{drug} with no further correction needed—though this assumes, notably, that an active treatment does not itself alter the magnitude of the placebo response, an assumption that is difficult to verify directly. This symmetry assumption may be particularly challenged in therapeutic areas where active treatment produces perceptible side effects that partially unmask assignment—potentially amplifying expectation-driven responses in the treated arm—or conversely where adverse effects dampen positive expectations and reduce the placebo contribution relative to the control arm. In an open-label patient-as-own-control trial, there is no concurrent control arm to perform this cancellation. Investigators observe a total within-patient change—$d_{\text{total}} = d_{\text{drug}} + \delta_p$—and must subtract the estimated placebo contribution at the analysis stage to determine d_{drug}. This post-trial subtraction is the interpretive step that the RCT performs automatically through its structure.

Importantly, the placebo effect does not enter the power calculation for either design; both the RCT and the own-control design are powered to detect d_{drug}, and the sample size ratio between them depends only on the within-subject correlation (ρ) [1]. The question is not whether the own-control design needs more patients to handle placebo—it does not—but whether the post-trial placebo subtraction is scientifically credible. That is the question this article addresses.

Decades of meta-analytic research have produced reliable estimates of placebo effect sizes across disease categories, endpoint types, and reporter modalities. Beyond simple subtraction, a rich literature in statistical methodology now offers sophisticated tools for decomposing observed treatment effects into drug and placebo components. These methods—including temporal trajectory modeling, concordance analysis, Bayesian approaches, mixture modeling, machine learning prediction, and causal inference frameworks—provide a level of analytic rigor that goes well beyond subtracting an estimate from a meta-analysis. The availability and credibility of these tools is the central focus of this article.

2 | Defining the Placebo Effect and Related Biases

For readers who may encounter these terms in regulatory discussions or trial design consultations, we define the placebo effect and the family of related biases that can influence clinical trial outcomes. The term "placebo effect" often encompasses distinct phenomena with varying magnitudes, mechanisms, and trial design implications.

2.1 | The Placebo Effect

The placebo effect refers to a genuine physiological or psychological improvement in a patient's condition that results not from the pharmacological action of a treatment but from the patient's expectations, beliefs, and the context of the therapeutic encounter [4, 5]. Neuroimaging and neurochemical studies have demonstrated that placebo responses involve measurable changes in brain activity, endogenous opioid release, dopamine signaling, and immune function [6]. The placebo effect is not imaginary—it produces real, measurable biological changes in certain domains. However, these changes are driven by the experience of receiving treatment rather than by the treatment's specific mechanism of action.

As discussed in the Introduction, the DBPCRCT handles the placebo effect through structural cancellation: both arms experience similar placebo responses, and the between-group difference isolates the true drug effect. In an open-label own-control trial, the placebo contribution must instead be estimated and removed analytically. Understanding the magnitude and characteristics of each bias component is therefore essential for designing credible post-trial corrections.

2.2 | Related Biases in Clinical Measurement

Several additional phenomena are often conflated with the placebo effect but are mechanistically distinct:

17528062, 2026, 5, Downloaded from https://ascpt.onlinelibrary.wiley.com/doi/10.1111/cts.70550 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

2.3 | Regression to the Mean

Patients are typically enrolled in trials during symptomatic flares or at their most severe. Subsequent measurements will, on average, trend toward the population mean regardless of any intervention. Similarly, investigators may unconsciously inflate subjective assessments (e.g., even the exact measurement on a manual BP cuff) to match enrollment criteria, thus subtly inflating severity at study entry. These statistical artifacts can mimic a treatment response in any design lacking a concurrent control that experiences the same regression [7].

2.4 | The Hawthorne Effect

Patients who know they are being observed—through frequent clinic visits, laboratory monitoring, and investigator attention—may modify their behavior or report improved symptoms. This is particularly relevant in trials with subjective endpoints and may affect both open-label and blinded studies [8].

2.5 | Effort Dependence

Patients who know they are being given the experimental treatment may "try harder" on effort-dependent measures such as forced vital capacity (FVC), forced expiratory volume (FEV1), 6-min walk test (6MWT) distance, or timed ambulation assessments. This concern is not merely theoretical: the U.S. Food and Drug Administration (FDA) has cited effort dependence as a basis for questioning efficacy findings in rare disease drug applications, particularly when primary endpoints rely on volitional physical performance measures. In neuromuscular diseases such as Duchenne muscular dystrophy and spinal muscular atrophy, effort-dependent endpoints such as the North Star Ambulatory Assessment (NSAA) and pulmonary function tests are frequently the most clinically relevant outcomes available, creating a tension between regulatory skepticism of effort-dependent measures and clinical necessity. Effort dependence overlaps with both the Hawthorne effect and expectation bias, but it is mechanistically distinct in that it reflects a real change in patient behavior—increased motivation and exertion—rather than altered perception of symptoms. Importantly, for the patient-as-own-control design, effort dependence can be managed through design-level strategies including blinded assessor protocols, standardized encouragement scripts, and objective anchoring measures (e.g., respiratory gas exchange during walk tests) that distinguish genuine physiological improvement from motivational effort increases.

2.6 | Expectation Bias

In open-label trials, patients know they are receiving the active treatment. This knowledge may lead to optimistic reporting of subjective symptoms, particularly for patient-reported outcomes (PROs) such as pain scales, fatigue inventories, and quality-of-life questionnaires [9].

2.7 | Observer (Assessor) Bias

When clinicians or caregivers know the treatment assignment, their outcome assessments may be unconsciously influenced toward reporting improvement. This is distinct from the patient's own experience and can affect clinician-rated scales, physical examination findings, and even the interpretation of borderline results [10].

2.8 | Natural Disease Fluctuation

Many diseases, including some rare diseases, have waxing and waning courses. Improvement observed during a treatment period may reflect the natural disease cycle rather than a treatment effect or placebo response.

In practice, what is loosely termed the "placebo effect" in open-label trials encompasses all of these phenomena. For the purposes of trial design, the critical question is not which mechanism predominates but rather: *what is the total magnitude of non-pharmacological improvement, and how can it be accounted for quantitatively?*

3 | Measuring the Placebo Effect: Evidence From Meta-Analyses

The most rigorous approach to estimating placebo effect magnitude comes from meta-analyses comparing placebo-treated patients to untreated controls—studies that include a true "no treatment" arm alongside the placebo arm. This design isolates the placebo response from natural disease course and regression to the mean. The standardized mean difference (SMD, or Cohen's *d*) [11] between placebo and no-treatment groups provides a quantitative estimate of placebo effect size that can be applied in post-trial interpretation.

3.1 | Landmark Meta-Analyses of Placebo Effects

The foundational work of Hróbjartsson and Gøtzsche, first published in 2001 and updated in 2010, analyzed 202 trials that included both placebo and no-treatment arms [4, 5]. Their findings established several key principles. Across all outcome types, the overall placebo effect was modest, with an SMD of −0.23 (95% confidence interval [CI]: −0.28 to −0.17). However, this aggregate figure masks important heterogeneity by endpoint type. For continuous subjective endpoints such as pain, fatigue, and quality of life (QoL), the SMD was −0.36 (95% CI: −0.47 to −0.25)—a small-to-moderate effect. For objective continuous endpoints such as biomarker levels and physiological measurements, the SMD was only −0.12 (95% CI: −0.27 to 0.03), which did not reach statistical significance. For binary outcomes such as survival or event occurrence, there was no significant placebo effect (relative risk [RR] 0.95, 95% CI: 0.88–1.02).

Fendel et al. [12] conducted the largest meta-analysis of open-label placebo trials to date, encompassing 63 trials with 4554 participants. They reported an overall SMD of 0.35 (95% CI:

17528062, 2026, 5, Downloaded from https://ascpt.onlinelibrary.wiley.com/doi/10.1111/cts.70550 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

0.26–0.44), with larger effects for self-reported outcomes (SMD = 0.39) compared to objective measures (SMD = 0.09). Clinical populations showed larger effects (SMD = 0.47) than non-clinical samples.

For neurological conditions, Mestre et al. in comprehensive reviews of placebo effects in Parkinson's disease found that clinician-rated motor scores (Unified Parkinson's Disease Rating Scale [UPDRS]) show small-to-moderate placebo responses, while objective biomarkers such as dopamine transporter (DAT) imaging showed minimal placebo response, consistent with the objective endpoint findings above [13]. In rare metabolic diseases, where objective biomarker endpoints predominate, the available evidence indicates placebo effects on biochemical measures are consistently small (SMD 0.02–0.09), consistent with the objective-endpoint findings from the larger meta-analyses described above, though direct meta-analytic data from rare disease placebo arms remain limited [5, 12]. Brooks et al. [14] have further argued that rare disease trials can be expanded by grouping conditions with shared molecular etiologies, an approach that inherently emphasizes objective molecular endpoints where placebo effects are minimal.

Two recent studies further sharpen the objective–subjective distinction. Bschor et al. [15], in the largest meta-analysis of placebo outcomes across psychiatric disorders, found substantial placebo effects on clinician- and self-rated outcomes (SMD 0.30–0.72) across nine conditions, reinforcing the vulnerability of subjective and psychological endpoints. Botvinik-Nezer et al. [16], using neuroimaging, demonstrated that placebo treatment affected brain systems related to affective and cognitive processing (SMD ≈ 0.30) but had no significant effect on objective nociceptive pain pathways ($d = 0.08$), providing a neurobiological basis for the endpoint-type distinction that runs throughout this article.

3.2 | Summary of Placebo Effect Sizes Across Meta-Analyses

Table 1 summarizes placebo effect estimates from key meta-analyses, organized by population and endpoint type.

4 | Subjective Versus Objective Endpoints: A Critical Distinction

The most consistent meta-analytic finding is that placebo effects are strongly dependent on endpoint type—a distinction with direct implications for rare disease trial design, where endpoint selection is often the most consequential sponsor decision.

4.1 | Objective Endpoints

For objective, biologically anchored endpoints, placebo effects are consistently small (SMD < 0.10) and frequently indistinguishable from zero [4, 5, 12]. This is expected: expectation and belief cannot meaningfully alter phenylalanine concentrations, dystrophin expression, or hemoglobin levels. Approximately 72% of rare diseases have identified genetic origins [23], and the majority of approved rare disease therapies use objective primary endpoints. For trials using these endpoints, the post-trial placebo correction is minimal and any imprecision in the estimate is inconsequential.

4.2 | Subjective Endpoints

For subjective endpoints—patient-reported pain, fatigue, quality of life, and symptom severity—placebo effects are meaningful, with SMDs typically ranging from 0.20 to 0.50 [4, 5, 12]. This does not mean these endpoints should be avoided; many rare diseases produce symptoms that are best captured by subjective measures, and regulatory agencies appropriately value patient-reported outcomes. Rather, when subjective endpoints are used in open-label designs, the expected placebo contribution should be quantified and incorporated into the analysis plan a priori.

4.3 | Mixed Endpoint Strategies

An increasingly common and recommended approach in rare disease trials is to pair objective primary endpoints with subjective secondary endpoints, or vice versa. When objective and subjective measures move in the same direction and magnitude, the convergent evidence strengthens the case for a true treatment effect. When only the subjective measure improves while the objective measure does not, placebo or bias effects should be considered. This concordance analysis provides internal validation that is unavailable in single-endpoint designs.

5 | Susceptibility by Endpoint Type and Reporter: A Ranked Framework

To guide endpoint selection and statistical planning, Table 2 ranks common endpoint categories by their estimated susceptibility to placebo effects, from lowest (objective, instrument-determined) to highest (subjective, patient-reported psychological). This ranking synthesizes the meta-analytic evidence and provides an a priori framework for the degree of post-trial correction that may be needed.

5.1 | Reporter-Specific Considerations

The identity of the reporter adds an additional layer of variability. In pediatric rare diseases—which constitute a large proportion of genetic rare conditions—outcomes are frequently reported by parents or caregivers rather than patients. Eiser and Varni [24] and Upton et al. [25] demonstrated that parent-reported quality of life consistently exceeds child-reported measures, with discordance increasing for subjective domains. In conditions where the child cannot self-report, this effect is difficult to disentangle from true treatment benefit without objective anchoring endpoints.

Clinician assessments are also not immune to bias. Ahmadzad-Asl et al. [19] found that clinician-rated improvements in psychiatric trials exceeded self-reported improvements, suggesting that observer bias may amplify rather than temper apparent

17528062, 2026, 5, Downloaded from https://ascpt.onlinelibrary.wiley.com/doi/10.1111/cts.70550 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 1 | Placebo effect sizes (standardized mean differences) from major meta-analyses.

Study	Population	Outcome type	Subjective SMD	Objective SMD	Temporal notes
Hróbjartsson and Gøtzsche (2010) [4, 5]	Mixed clinical (202 trials)	Pain, fatigue, QoL	−0.36 (−0.47 to −0.25)	−0.12 (NS)	Stronger early; decays over weeks
Fendel et al. (2025) [12]	Clinical & non-clinical (63 trials)	Self-reported symptoms	0.39 (0.30–0.49)	0.09	Fades in trials > 12 weeks
Von Wernsdorff et al. (2021) [17]	Clinical (IBS, depression; 11 trials)	Self-reported symptoms	0.72 (0.39–1.05)	Not assessed	Peaks in first month; high heterogeneity
Mestre et al. (2018) [13]; Lidstone et al. (2010) [18]	Parkinson's disease (multiple trials)	Motor scores (clinician)	~0.23–0.39 (UPDRS; per review)	0.05 (NS)	Diminishes in open-label extensions
Hróbjartsson (2010); Fendel (2025) [5, 12]	Metabolic rare diseases (applied)	Biomarkers, molecular endpoints	≤0.30 (PROs)	0.02–0.09	General meta-analytic objective data applied to rare disease biomarker endpoints
Ahmadzad-Asl et al. (2022) [19]	Psychiatric (depression)	Clinician-rated scales	0.27 (0.14–0.41)	Not assessed	Clinician > self-report early
Bschor et al. (2024) [15]	Psychiatric (9 disorders)	Clinician- & self-rated	0.30–0.72	Not assessed	Substantial across disorders; varies by condition
Botvinik-Nezer et al. (2024) [16]	Pain (neuroimaging)	Affective/cognitive vs. nociceptive	0.30 (affective)	0.08 (nociceptive)	Objective nociceptive processes unaffected
Vase et al. (2009) [20]	Pain (198 trials)	Pain intensity (VAS)	0.15–0.57	< 0.20	Decays in open-label extensions
Chen et al. (2017) [21]	Fibromyalgia (229 trials)	Pain, fatigue	0.53 (blinded); ~0.25 (open)	Not assessed	Halved in open-label settings

Note: SMD signs vary by convention across meta-analyses; negative values indicate improvement in some scales (e.g., pain), positive values in others (e.g., function). Absolute values are reported here for comparability; readers should consult the original publications for directionality. These SMD estimates are derived predominantly from common-disease populations. Rare disease–specific data (e.g., Ghadessi et al. [22]) suggest similar patterns for objective endpoints, though direct meta-analytic data from rare disease placebo arms remain limited. All SMD estimates represent means across heterogeneous trials; confidence intervals are wide in several cells (e.g., Von Wernsdorff et al.: 95% CI 0.39–1.05; Hróbjartsson and Gøtzsche objective: 95% CI −0.27 to 0.03), and these figures should be interpreted as approximate summaries rather than precise priors.

Abbreviations: CI, confidence interval; IBS, irritable bowel syndrome; NS, not statistically significant; OLP, open-label placebo; PD, Parkinson's disease; PRO, patient-reported outcome; QoL, quality of life; SMD, standardized mean difference; VAS, visual analog scale.

17528062, 2026, 5, Downloaded from https://ascpt.onlinelibrary.wiley.com/doi/10.1111/cts.70550 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 2 | Ranked endpoint susceptibility to placebo effects in open-label trials.

Rank	Endpoint category	Examples	Estimated placebo SMD	Rationale/Evidence
1	Survival/time-to-event	Overall survival, event-free survival	0.00–0.01	No significant placebo effect on hard survival endpoints in any meta-analysis (Hróbjartsson 2001, 2010 [4, 5]: RR 0.95, NS)
2	Laboratory biomarkers (direct)	Enzyme activity, serum protein, metabolite levels	0.02–0.05	Biochemical measurements unaffected by expectation (Hróbjartsson and Gøtzsche 2010 [5]: objective SMD −0.12, NS; Fendel et al. 2025 [12]: objective SMD 0.09)
3	Molecular/genetic markers	Dystrophin expression, gene expression panels	0.00–0.03	Gene expression not modifiable by psychological state; measurement fully automated
4	Imaging-based volumetric/ quantitative	MRI volumetrics, CT bone density, DEXA	0.03–0.08	Quantitative imaging resistant to bias; Mestre et al. 2018 [13]: DAT imaging minimal placebo response in PD; Botvinik-Nezer et al. 2024 [16]: nociceptive $d = 0.08$
5	Physiological/functional (objective)	FVC, 6MWT distance, cardiac output	0.05–0.12	Some effort-dependent component; partial placebo susceptibility but anchored in physiology
6	Clinician-rated composite scales	UPDRS, NSAA, ALSFRS-R	0.15–0.30	Observer bias in open-label settings (Ahmadzad-Asl et al. 2022 [19]: clinician SMD 0.27); mitigable by blinding assessors
7	Caregiver/parent-reported outcomes	PedsQL (parent proxy), Vineland (parent)	0.20–0.40	Caregiver hope and investment inflate reports; Eiser and Varni 2011 [24]: parent–child discordance in QoL
8	Patient-reported outcomes (physical)	Pain VAS, fatigue scales, symptom diaries	0.25–0.50	Core domain of placebo response; Vase et al. 2009 [20]: pain SMD 0.15–0.57; Chen et al. 2017 [21]: fibromyalgia pain ES 0.53; Botvinik-Nezer et al. 2024 [16]: affective SMD 0.30
9	Patient-reported outcomes (psychological)	Depression scales, anxiety inventories, global impression	0.30–0.72	Highest placebo susceptibility; Von Wernsdorff et al. 2021 [17]: OLP SMD 0.72; Bschor et al. 2024 [15]: SMD 0.30–0.72 across 9 psychiatric disorders

Note: Green shading = objective endpoints (minimal placebo concern). Yellow = mixed. Orange = subjective (requires analytical management). SMD ranges are approximate summaries derived from the meta-analyses cited in Table 1; they reflect central estimates across heterogeneous trials and should not be treated as precise priors. Confidence intervals are wide for several endpoint categories, particularly those in the subjective domain. Sponsors should derive disease- and endpoint-specific estimates from the most closely matched available literature.

Abbreviations: 6MWT, 6-minute walk test; ALSFRS-*R*, Amyotrophic Lateral Sclerosis Functional Rating Scale–Revised; CT, computed tomography; DAT, dopamine transporter; DEXA, dual-energy X-ray absorptiometry; FVC, forced vital capacity; MRI, magnetic resonance imaging; NSAA, North Star Ambulatory Assessment; OLP, open-label placebo; PD, Parkinson's disease; PedsQL, Pediatric Quality of Life Inventory; QoL, quality of life; RR, relative risk; UPDRS, Unified Parkinson's Disease Rating Scale; VAS, visual analog scale.

17528062, 2026, 5, Downloaded from https://ascpt.onlinelibrary.wiley.com/doi/10.1111/cts.70550 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

improvement. In open-label rare disease trials, clinicians who have advocated for and administered a novel therapy may be susceptible to confirmation bias. Blinded outcome assessment—in which the clinician performing the evaluation is unaware of the treatment timeline—can mitigate this concern even in otherwise open-label designs.

6 | Temporal Dynamics: Does the Placebo Effect Fade?

A critical question for trial design is whether placebo effects are sustained or transient. If placebo responses fade over time while true treatment effects persist, then trial designs with multiple assessment points can use this divergence to help separate signal from noise.

6.1 | Evidence for Placebo Decay

Placebo effects peak early and diminish over time. Fendel et al. [12] found that trials lasting longer than 12 weeks showed smaller placebo effects than shorter trials. In major depressive disorder (MDD) trials, Meister et al. [26] reported that placebo effect sizes were largest in the first weeks of treatment and declined with continued follow-up. Mestre et al. [13] observed that placebo responses on motor scores in Parkinson's trials diminished during open-label extension phases. In pain research, Vase et al. [20] documented placebo effect decay in open-label extensions.

The temporal profile of placebo decay can be approximated by an exponential function:

$$\delta_p(t) = \delta_{p0} \times e^{-\lambda t}$$

where δ_{p0} is the initial placebo effect size and λ is the decay constant (t is time in weeks). The non-zero value at $t = 0$ reflects the well-documented anticipatory placebo response: enrollment and the expectation of receiving treatment generate an immediate improvement before the first dose is administered [6]. The range $\lambda = 0.05$–0.15 per week corresponds to a placebo half-life of approximately 5–14 weeks, derived from the finding of Fendel et al. [12] that trials lasting longer than 12 weeks show substantially smaller placebo effects, and from Meister et al. [26] and Vase et al. [20], who documented the most rapid decay in the first 4–8 weeks of open-label follow-up. Alternative temporal models—such as the Bateman function, which uses separate onset and offset rate constants—may better characterize diseases where the placebo response builds gradually before declining; the single-parameter exponential is used here for parsimony. For a typical $\delta_{p0} = 0.35$ and $\lambda = 0.10$, the placebo effect diminishes to approximately 0.21 by week 5, 0.13 by week 10, and 0.06 by week 20.

6.2 | Implications for Trial Design

Trials that collect outcome data at multiple time points have a significant advantage in managing placebo effects. When outcomes are assessed repeatedly—for example, at baseline, weeks 4, 8, 12, and 24—the time-course of improvement can be examined. A true pharmacological effect is expected to emerge at a time consistent with the drug's mechanism and persist or increase over time. A placebo-driven response, by contrast, typically appears immediately and decays. This temporal divergence is a powerful analytic tool, and we describe the statistical methods for exploiting it in the section on advanced analytical methods below. Figure 1 illustrates these contrasting trajectories schematically.

7 | Placebo Effects in Small Populations

A relevant question for rare disease trials is whether placebo effects behave differently in small samples. Hróbjartsson and Gøtzsche [5] observed that smaller trials tended to report larger placebo effects, but attributed this primarily to publication bias and methodological limitations rather than a true small-sample amplification. In rare disease specifically, Ghadessi et al. [22] analyzed historical controls in 45 FDA approvals using external comparators and found that placebo-like effects in objective endpoints were minimal (SMD < 0.10) regardless of sample size.

However, small samples are more vulnerable to the *variability* of placebo effects. In a trial of 200 patients, individual placebo responses average out; in a trial of 15 patients, a few strong placebo responders can meaningfully shift the group mean. This variability—rather than the mean placebo effect—is the primary statistical concern in small rare disease trials.

The patient-as-own-control design partially addresses this concern. Because each patient serves as their own baseline, the relevant comparison is within-patient change. A patient who shows a large placebo-driven improvement at an early timepoint but then regresses will demonstrate the temporal pattern described above, which can be modeled and accounted for. In a small parallel-group RCT, by contrast, the same placebo responders in the control arm would inflate the control mean and reduce power to detect a treatment effect—a problem that is difficult to detect and correct without additional analyses at small sample sizes.

8 | How the Two Designs Handle Placebo: A Conceptual Comparison

Having reviewed the evidence on placebo effect magnitudes, it is worth comparing directly how the DBPCRCT and the patient-as-own-control design each deal with placebo effects in post-trial analysis. The mathematical details of sample size calculations have been presented elsewhere [1]; here we focus on the interpretive logic.

8.1 | The RCT: Automatic Cancellation

In the RCT, both treatment and placebo arms experience the same placebo response. The between-group difference cancels the placebo component, yielding d_{drug} directly. No post-trial placebo estimation is needed. This is the RCT's principal interpretive advantage.

17528062, 2026, 5, Downloaded from https://ascpt.onlinelibrary.wiley.com/doi/10.1111/cts.70550 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

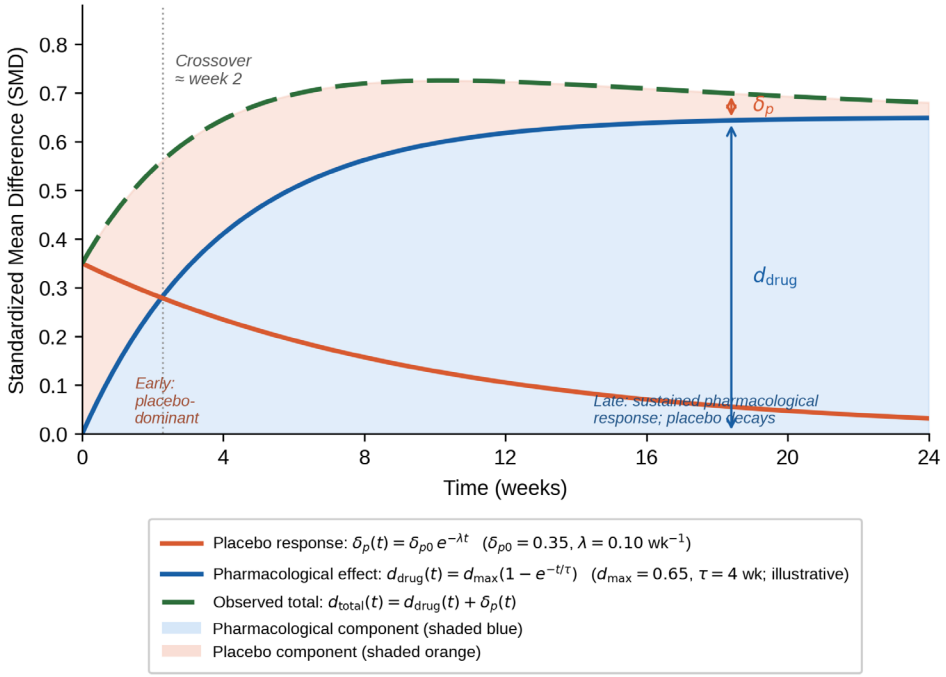


FIGURE 1 | Temporal separation of placebo and pharmacological responses in a patient-as-own-control trial. The placebo response (orange) is modeled as $\delta p(t) = \delta p0\ e - \lambda t$ with peak $\delta p0 = 0.35$ (mean subjective-endpoint SMD; Hróbjartsson and Gøtzsche [5], Fendel et al. [12]) and $\lambda = 0.10$ week−1 (half-life ≈ 7 weeks; derived from Fendel et al. [12], Meister et al. [26], Vase et al. [20]). The non-zero intercept at $t = 0$ reflects the anticipatory placebo response generated by enrollment and treatment expectation [6]. The pharmacological effect (blue) is modeled as $d$drug$(t) = d$max$(1 - e - t/\tau)$ with $d$max $= 0.65$ and $\tau = 4$ weeks (illustrative parameters only; not derived from a specific trial). The dashed green curve shows the observed total within-patient change, $d$total$(t) = d$drug$(t) + \delta p(t)$. Shaded regions decompose the total into its pharmacological (blue) and placebo (orange) components. The temporal divergence between the early-peaking, decaying placebo response and the rising, sustained pharmacological effect provides the analytical basis for temporal trajectory modeling as a placebo management strategy. Vertical brackets at week 18 illustrate the decomposition of $d$total at a late time point.

However, this elegant cancellation mechanism assumes that the placebo effect is equal in magnitude across both arms—an assumption that holds reliably only in large samples. In the small sample sizes characteristic of rare disease trials (often $N = 15$–50 total), placebo responses are subject to considerable sampling variability. A few strong placebo responders randomly allocated to the control arm can inflate the control group mean, narrowing the observed between-group difference and reducing the apparent treatment effect. Conversely, strong placebo responders in the treatment arm can inflate the treatment group mean, creating the appearance of a larger drug effect than truly exists. In either case, the assumed symmetry of placebo cancellation breaks down. This asymmetry is difficult to detect and correct without additional analyses in the standard RCT framework, because the trial provides no built-in mechanism for identifying individual placebo responders or estimating the magnitude of the imbalance. Moreover, effort dependence—a concern the FDA has raised in reviewing rare disease applications—can compound this problem: in a small blinded RCT where patients may suspect their treatment assignment based on perceived side effects or lack thereof, differential effort on performance-based endpoints can further distort between-arm comparisons. The result is that the RCT's structural advantage in handling placebo, while genuine in principle, is substantially degraded in the very populations where it is most needed.

The cost is also in patients. The RCT requires approximately 19 times more patients than the own-control design at $\rho = 0.90$ for a continuous endpoint with equal variance, 80% power, and $\alpha = 0.05$ (two-tailed), as derived in table 3 of the companion article [1]. For most rare diseases, this requirement exceeds the available patient population, rendering the RCT's interpretive elegance moot.

8.2 | The Own-Control: Estimation and Subtraction

In the own-control design, the observed within-patient change includes both the drug effect and the placebo contribution. The investigator subtracts the estimated δ_p from the observed d_{total} to recover d_{drug}. This introduces reliance on an external estimate. Gupta et al. [27] provide formal recommendations for clinical trial design in patients as their own control in rare diseases, including strategies for managing this estimation requirement.

Two observations temper this concern. First, for objective endpoints (the majority of primary endpoints in rare disease), the placebo effect is consistently small (SMD < 0.10) and in most meta-analyses statistically indistinguishable from zero. An imprecise estimate of a consistently small quantity has negligible impact. Second, for subjective endpoints where the placebo effect is larger, the uncertainty can be managed through the suite of advanced statistical methods described in the following section. These methods go well beyond simple

17528062, 2026, 5, Downloaded from https://ascpt.onlinelibrary.wiley.com/doi/10.1111/cts.70550 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

subtraction and provide multiple, complementary lines of evidence for decomposing d_{total} into its drug and placebo components.

A third observation is often overlooked but deserves emphasis: for diseases with patient populations too small to support an adequately powered RCT, the patient-as-own-control design may represent not merely a convenient alternative but the optimal scientifically rigorous path to evaluating therapeutic efficacy. An underpowered RCT—one that enrolls 30 patients when 125 are needed—does not provide a "less certain" answer; it provides an answer that is statistically unreliable and prone to both false negatives (missing real effects) and the small-sample placebo distortions described above. The placebo concerns associated with open-label own-control designs are manageable through the analytical methods described in this article—temporal trajectory modeling, concordance analysis, Bayesian adjustment, and blinded outcome assessment—provided that these strategies are incorporated through careful forethought at the design stage. By contrast, the methodological compromises forced by underpowered RCTs are not correctable after the fact. For the majority of rare and ultra-rare diseases, a well-designed own-control study with pre-specified placebo management yields more credible evidence than an underpowered RCT that superficially resembles the gold standard but lacks the statistical power to deliver on its promise.

8.3 | The Trade-Off

Both designs ultimately estimate the same quantity: d_{drug}. The RCT does so without requiring a placebo estimate; the own-control design does so with an approximately 19-fold reduction in patients at $\rho = 0.90$ (table 3, Summar and Woodcock [1]). For a rare disease with 50 patients worldwide, the choice is between an RCT that cannot enroll and an own-control trial that can enroll but requires a credible post-trial placebo correction. The question is not which design is theoretically preferable in the abstract, but which is feasible in the reality of rare disease. The remainder of this article demonstrates that the post-trial correction tools available are rigorous enough to support the own-control design. Figure 2 illustrates the structural logic of each design side by side.

A simple numerical illustration clarifies the practical consequence of placebo responder imbalance in small-sample RCTs. Consider a parallel-group RCT of $N = 36$ patients randomized 1:1 ($n = 18$ per arm) testing a drug with true effect $d_{drug} = 0.5$. Assume a placebo responder rate of 20% ($\delta_p = 0.5$ SD in responders, 0 in non-responders). Under balanced allocation, each arm receives approximately 3–4 placebo responders, and the between-arm difference reflects $d_{drug} \approx 0.5$ as intended. Under asymmetric allocation—a plausible outcome by chance at $N = 36$–6 of the 7 placebo responders fall in the control arm. The control arm mean rises by approximately 0.17 SD ($6 \times 0.5/18$), while the treatment arm mean is unaffected. The observed between-arm difference narrows to approximately 0.33, falling below the threshold of clinical meaningfulness and failing to reach statistical significance at $\alpha = 0.05$ with this sample size. The drug is declared ineffective despite producing its true effect. By contrast, in an own-control design enrolling 8 patients (the number required at $d = 0.5$, $\rho = 0.90$), each patient serves as their own baseline. Placebo-driven improvement appears early in each patient's trajectory and decays, while the sustained pharmacological response persists; temporal trajectory analysis separates the two components. No asymmetric allocation between arms is possible because there are no arms. This example illustrates that the RCT's structural cancellation of placebo, while valid in expectation, fails in the small samples characteristic of rare disease, and that this failure mode is not detectable or correctable post hoc without additional analyses.

9 | The Additivity Assumption and Its Limits

The framework described above—$d_{total} = d_{drug} + \delta_p$—rests on the assumption that drug and placebo effects are additive: the total observed change equals the sum of the two independent components. This additivity assumption is standard in clinical trial analysis, but it has been questioned.

Enck et al. [28] raised the question of whether drug and placebo effects interact rather than simply summing. If a drug produces genuine physiological improvement, the patient may experience that improvement, develop additional positive expectations, and thereby generate a larger placebo response than would occur in a placebo-only condition. Conversely, side effects from an active drug (fatigue, nausea) might dampen the positive expectations that drive placebo responses. Under either scenario, the placebo effect in a drug-treated patient may differ from the placebo effect measured in a placebo-only arm.

Curković et al. [29] explored this issue in the context of antidepressant trials, noting that some analyses attribute the majority of the observed clinical improvement to placebo-related mechanisms rather than pharmacological action, and that the interaction between drug and placebo components may not be cleanly separable. The concept of "active placebo" controls—agents that mimic side effects without therapeutic action—was proposed to test this interaction, and trials using active placebos have sometimes shown smaller drug-placebo differences than trials using inert placebos.

For rare disease trials, the practical implications of the additivity question are moderated by endpoint type. For objective endpoints, where the placebo contribution is consistently small (SMD < 0.10) and frequently indistinguishable from zero, any interaction with the drug effect is correspondingly minimal—the product of a small number and any multiplicative factor remains small. The additivity assumption is therefore robust for the endpoint types most commonly used in rare disease.

For subjective endpoints, the additivity assumption deserves more caution. Recent meta-analytic evidence suggests that drug–placebo interactions may be more pronounced in subjective domains: Bschor et al. [15] reported that placebo effect magnitudes varied substantially across psychiatric conditions (SMD 0.30–0.72) and were modulated by trial design features such as the probability of receiving active treatment, suggesting

17528062, 2026, 5, Downloaded from https://ascpt.onlinelibrary.wiley.com/doi/10.1111/cts.70550 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

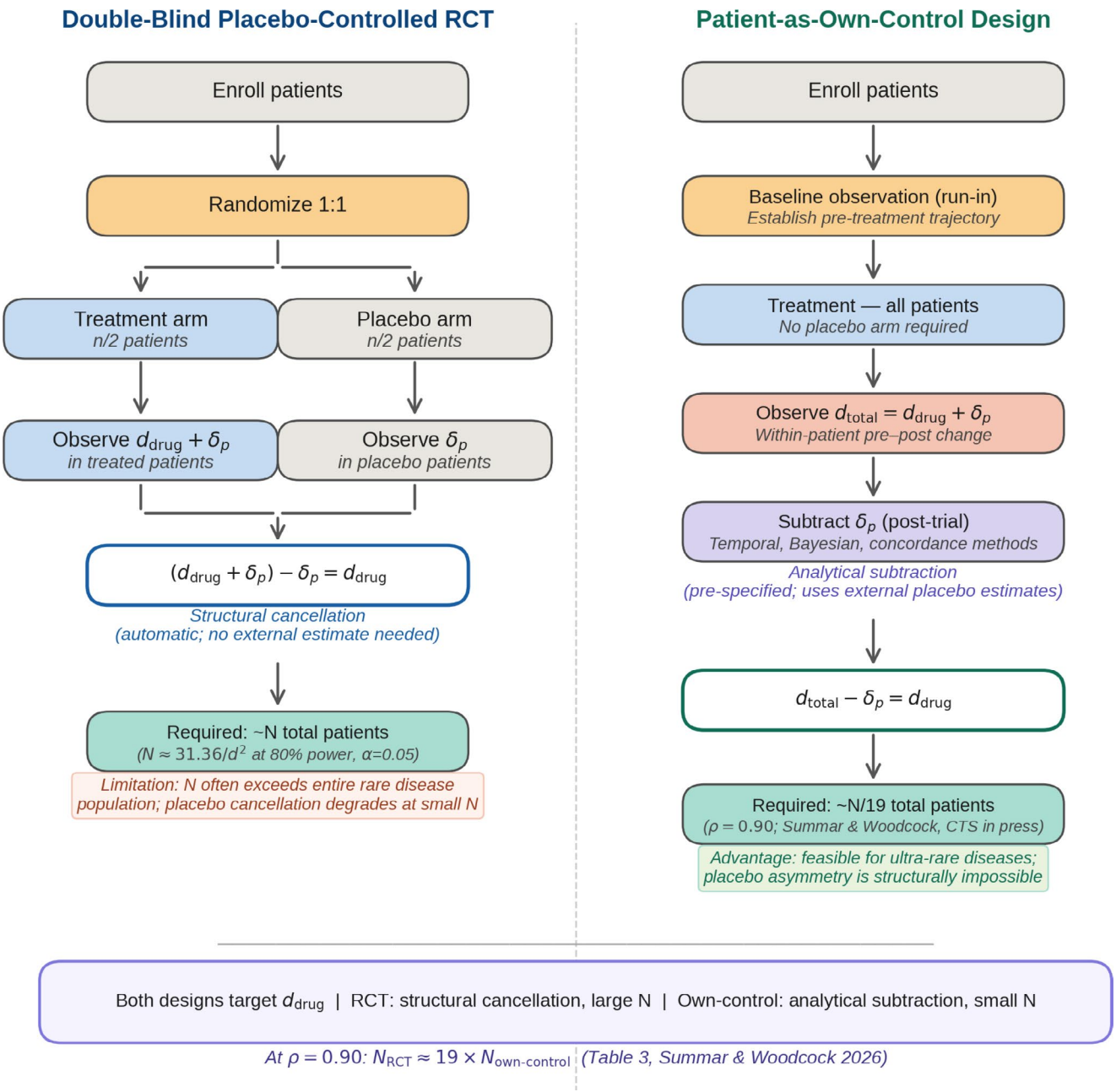


FIGURE 2 | Design comparison: Double-blind placebo-controlled RCT versus patient-as-own-control design. Both designs target the true drug effect ddrug but achieve this through different mechanisms. In the RCT (left), randomization to treatment and placebo arms enables structural cancellation of the placebo contribution when the between-arm difference is computed, requiring no external placebo estimate; however, total sample size $N \approx 31.36/d2$ routinely exceeds the available rare disease population, and placebo cancellation degrades at small N due to asymmetric responder allocation. In the patient-as-own-control design (right), all patients receive treatment following a baseline observation period; the observed total change $d\text{total} = d\text{drug} + \delta p$ is decomposed analytically post-trial using temporal trajectory modeling, Bayesian priors, or concordance analysis. At within-subject correlation $\rho = 0.90$, the own-control design requires approximately 1/19th the patients of an RCT at equivalent power (Table 3, Summar and Woodcock [1]).

context-dependent interactions rather than simple additivity. We recommend that investigators treat the back-end subtraction as an approximation and use sensitivity analyses that explore a range of δ_p values, including values both above and below the meta-analytic estimate. The advanced methods described below—particularly growth mixture modeling and Bayesian approaches—can accommodate non-additive interaction models, providing more nuanced decompositions when the additivity assumption is explicitly questioned.

10 | Advanced Statistical Methods for Post-Trial Placebo Adjustment

The discussion of placebo in open-label trials is sometimes framed as if the only tool available were a simple subtraction: observe d_{total}, subtract an estimate of δ_p, report d_{drug}. In reality, decades of research in placebo science, clinical trial methodology, and causal inference have produced a rich toolkit of statistical approaches that go well beyond simple subtraction. These

17528062, 2026, 5, Downloaded from https://ascpt.onlinelibrary.wiley.com/doi/10.1111/cts.70550 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 3 | Applicability of post-trial placebo adjustment methods by endpoint type.

Analytical method	Objective endpoints	Subjective endpoints	Data requirements	Complexity	Example rare disease application
Sensitivity analysis (subtraction)	Sufficient alone	Useful as baseline	Meta-analytic δp	Low	Enzyme replacement in lysosomal storage diseases
Temporal trajectory modeling	Supportive	Strong evidence	Serial assessments (≥3 timepoints)	Moderate	Neuromuscular diseases with serial 6MWT
Concordance validation	Reference standard	Strong validation	Paired obj + subj endpoints	Low	Pompe disease (enzyme levels + functional measures)
Bayesian informative priors	Optional	Recommended	Prior distribution from literature	Moderate	Ultra-rare metabolic diseases with limited trial data
Growth mixture modeling	Rarely needed	Powerful	Longitudinal data, $N \geq 20$–30	High	Neurodegenerative diseases with heterogeneous progression
Latent class/regression	Rarely needed	Powerful	Baseline covariates, longitudinal	High	Genetic epilepsies with variable phenotypes
ML placebo prediction	Rarely needed	Promising	Historical trial training data	High	Diseases with prior trial databases (e.g., DMD)
Causal inference (PSM, IPTW, etc.)	Useful with ext. controls	Useful with ext. controls	External control dataset	Moderate–High	Diseases with established registries (e.g., cystic fibrosis)
SPCD/enrichment	Not applicable	Design-level solution	Blinded run-in phase	Moderate	Chronic fatigue or pain endpoints in rare diseases
Extended run-in observation	Strong baseline calibration	Design-level solution; addresses effort dependence	Stable or slowly progressive disease; serial assessments	Low–Moderate	Slowly progressive neuromuscular diseases (SMA, DMD)
Blinded outcome assessment	Standard practice	Essential for clinician scales	Blinded assessors	Low	Any rare disease with clinician-rated functional scales

Abbreviations: AIPW, augmented inverse probability weighting; Ext., external; GMM, growth mixture modeling; IPTW, inverse probability of treatment weighting; LCA, latent class analysis; ML, machine learning; PSM, propensity score matching; SPCD, sequential parallel comparison design; TMLE, targeted maximum likelihood estimation.

17528062, 2026, 5, Downloaded from https://ascpt.onlinelibrary.wiley.com/doi/10.1111/cts.70550 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

methods provide multiple complementary lines of evidence for separating drug from placebo effects, and their combined application can achieve a level of analytic rigor that substantially addresses the interpretive concerns associated with open-label designs. We present them here organized by increasing sophistication.

10.1 | Pre-Specified Sensitivity Analysis With Placebo Subtraction

The simplest approach is to subtract a range of estimated placebo effects from the observed treatment effect in a pre-specified sensitivity analysis. If the observed within-patient improvement is $d_{obs} = 0.60$, the statistical analysis plan specifies testing against δ_p values of 0.10, 0.20, 0.30, and 0.40. If the placebo-adjusted effect remains statistically significant and clinically meaningful across this range, the treatment effect is robust. This approach is transparent and easy to communicate to regulatory reviewers but is limited by its reliance on aggregate estimates.

10.1.1 | When Adequate

For objective endpoints where $\delta_p < 0.10$, simple subtraction is more than sufficient. The correction is small, and imprecision in the estimate is inconsequential.

10.2 | Temporal Trajectory Modeling

When outcomes are assessed at multiple timepoints, the divergent time-courses of drug and placebo effects provide powerful evidence for decomposition. As described above, placebo effects characteristically peak early—typically within the first 4–8 weeks—and then decay as the novelty and expectation that drive them diminish. True pharmacological effects, by contrast, emerge on a timescale consistent with the drug's mechanism of action and typically persist or increase with continued treatment. This temporal divergence can be formally modeled using linear mixed-effects models (LMMs) that include both a treatment-time interaction and an estimated placebo decay term. The model takes the following form:

$$Y_{ij} = \beta_0 + \beta_1(\text{Treatment}_i) + \beta_2(\text{Time}_j) + \beta_3(\text{Treatment} \times \text{Time}) - \delta_p(t_j) + b_i + \varepsilon_{ij}$$

Each variable in this equation serves a specific role. Y_{ij} is the observed outcome for patient i at timepoint j—the primary endpoint measurement at each scheduled assessment. The subscript i indexes individual patients ($i = 1, 2, ..., N$) and j indexes the sequential measurement occasions (e.g., baseline, week 4, week 8, week 12).

β_0 is the intercept, representing the average baseline outcome across all patients at time zero (i.e., before any treatment or placebo effect has occurred). β_1 captures the immediate average difference at baseline between the treatment group and the control condition; in a well-designed study this should be near zero. In the patient-as-own-control application specifically, Treatment ~ i ~ indexes the post-treatment phase (coded 1) versus the pre-treatment phase (coded 0) within the same patient, rather than between-group assignment—the "control condition" is each patient's own pre-treatment period. β_2 represents the average rate of change over time that is common to all patients regardless of treatment assignment—this accounts for natural disease progression, practice effects from repeated testing, or other time-related trends.

The term of primary interest is β_3, the treatment-by-time interaction. This coefficient captures the *additional* rate of change attributable to the drug beyond the background time trend. A statistically significant β_3 indicates that treated patients improve (or decline more slowly) at a different rate than would be expected from time effects alone, and this differential rate is the core evidence for pharmacological efficacy.

The term $\delta_p(t_j)$ is the estimated placebo decay function, which models the time-varying contribution of placebo-related improvement. This term incorporates the exponential decay described in the preceding section (e.g., $\delta_p(t) = \delta_{p,\max} \times e^{-\lambda t}$, where $\delta_{p,\max}$ is the peak placebo effect and λ is the decay rate). By explicitly including this term, the model separates early improvements that fade—the hallmark of placebo response—from sustained improvements that persist, which are more likely pharmacological in origin.

The random effect b_i accounts for the fact that patients differ from one another in their baseline severity, rate of progression, and responsiveness. Including b_i allows each patient's trajectory to deviate from the population average, which is essential in rare disease populations where inter-patient heterogeneity is often substantial. Finally, ε_{ij} is the residual error term, representing the random, unexplained variation in each measurement occasion that is not captured by any of the systematic terms in the model.

The practical interpretation is straightforward: if a treated patient's improvement is sustained beyond the 8–12 week window where placebo effects typically dissipate, and if β_3 remains significant after accounting for the placebo decay term, this pattern constitutes strong evidence that the observed benefit is pharmacological rather than placebo-mediated. The model thus leverages the distinct temporal signatures of drug and placebo effects to disentangle them, even in single-arm designs where no concurrent placebo group is available for direct comparison.

10.3 | Concordance Between Objective and Subjective Endpoints

When both objective and subjective endpoints are included in a trial, their concordance provides powerful internal validation. Objective endpoints are minimally susceptible to placebo (Table 2). If the objective endpoint improves, the treatment is producing genuine biological change, and concurrent subjective improvement is more likely genuine. If only the subjective measure improves while the objective measure does not, placebo or bias effects become the leading explanation. This internal cross-validation is specific to the trial's own data and requires no external assumptions about placebo magnitude.

17528062, 2026, 5, Downloaded from https://ascpt.onlinelibrary.wiley.com/doi/10.1111/cts.70550 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

10.4 | Bayesian Informative Priors for Placebo Adjustment

Bayesian statistical methods offer a formal mechanism for incorporating prior knowledge about placebo effects directly into the statistical model [30]. In brief, a Bayesian analysis combines observed trial data with a "prior distribution"—a mathematical representation of what is already known from previous studies—to produce a "posterior distribution" that represents the updated estimate of the treatment effect. This approach is particularly well-suited to rare disease trials where sample sizes are small and external information is valuable. Hampson et al. [31] demonstrated that Bayesian designs can reduce required sample sizes by 20%–40% in rare disease settings while maintaining appropriate error control. Notably, the FDA's draft guidance on the use of Bayesian methodology in clinical trials endorses the use of informative priors for incorporating external data and subgroup analyses, providing direct regulatory support for this approach in rare disease settings [32].

In the context of placebo adjustment, the approach proceeds as follows. From published meta-analyses (Table 1), a prior distribution for the placebo effect is constructed—for example, $\delta_p \sim \text{Normal}(0.35,\ 0.10^2)$ for patient-reported pain. A separate prior for the drug effect is specified based on preclinical data. The observed within-patient change is modeled as the sum: $d_{\text{obs}} = \delta_{\text{drug}} + \delta_p + \varepsilon$. The Bayesian framework uses the trial data to update both priors simultaneously, producing posterior distributions for both the drug effect and the placebo contribution.

Critically, this approach does not require *knowing* the exact placebo effect—it accounts for *uncertainty* about the placebo effect through the width of the prior distribution. A wider prior (reflecting greater uncertainty) leads to a more conservative posterior estimate of the drug effect. This built-in conservatism makes Bayesian methods particularly appropriate when disease-specific placebo data are limited.

10.5 | Advanced Decomposition Methods: Mixture Models, Machine Learning, and Causal Inference

Beyond the methods described above, a family of more computationally intensive approaches can decompose observed treatment effects into drug and placebo components with greater granularity. These methods are particularly relevant for subjective endpoints in trials with sufficient longitudinal data, and we summarize them here: readers seeking implementation details are referred to the cited primary sources.

Growth mixture modeling (GMM) identifies latent subpopulations of patients with different response patterns—drug-only responders, placebo-only responders, always responders, and never responders—using longitudinal outcome data [33]. By estimating the probability of class membership for each patient, GMM isolates the pharmacological effect without requiring the additivity assumption. The method accommodates the between-patient heterogeneity characteristic of rare disease populations but requires a minimum of approximately 20–30 patients per class for reliable estimation, and model selection criteria (Bayesian Information Criterion, entropy) should be interpreted conservatively in small samples.

Latent class and latent regression models extend this framework by incorporating baseline covariates—disease severity, biomarker levels, demographic characteristics—to predict class membership and estimate a continuous spectrum of placebo susceptibility across patients [34, 35]. Three-step estimation approaches with BCH (Bolck–Croon–Hagenaars) correction methods for unbiased covariate estimation are preferred [36].

Machine learning–based placebo prediction trains algorithms on historical trial data to generate individual-level predicted placebo response scores, which are then subtracted from observed outcomes [37]. Published validation data report improvements in assay sensitivity of approximately 40% and power gains of approximately 14%, at a cost of 1%–3% of total per-patient trial costs. The principal constraint for rare diseases is the availability of training data; for ultra-rare diseases with no prior trial history, cross-disease models are an active area of development.

Causal inference methods for external controls, including propensity score matching, inverse probability of treatment weighting (IPTW), augmented inverse probability weighting (AIPW), targeted maximum likelihood estimation (TMLE), and G-computation, provide rigorous tools for estimating treatment effects when own-control trials are supplemented with historical or natural history control data [38–42]. These methods estimate the counterfactual—what would have happened to the treated patient without treatment—which encompasses both natural disease course and placebo effects. All require the "no unmeasured confounders" assumption, which is stronger in the external control setting than in the randomized setting, but when well-matched external controls are available through standardized registries and natural history studies, these methods provide a principled analytical framework.

For rare disease trials, these advanced methods are most likely to be employed for subjective endpoints in diseases with prior clinical trial or natural history data. For objective endpoints, the simpler methods described in sections 1–4 are typically sufficient, and the correction is minimal. Consultation with an experienced biostatistics team is recommended when implementing any of these approaches.

10.6 | Design-Level Approaches: Enrichment, Sequential Parallel Comparison, and Extended Run-In Observation

Although the focus of this article is on post-trial analytical methods, several design-level strategies deserve mention because they directly address placebo effects.

10.6.1 | Placebo Run-In Enrichment

Patients are given placebo during a pre-randomization period, and those who show a substantial placebo response are excluded from the active trial. This reduces placebo "noise" in the

17528062, 2026, 5, Downloaded from https://ascpt.onlinelibrary.wiley.com/doi/10.1111/cts.70550 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

treatment phase. While this approach requires blinding and is therefore not applicable to all open-label designs, it is feasible in crossover designs and blinded own-control studies [43].

10.6.2 | Sequential Parallel Comparison Design (SPCD)

Raman [42] describes SPCD as a two-stage design in which placebo non-responders from stage 1 are re-randomized in stage 2, enriching the sample for patients whose improvement is more likely to reflect drug effect. SPCD has shown particular promise in psychiatric and neurological trials where placebo response rates are high.

Accurate patient reporting training (APRT), in which patients are trained to report symptoms more consistently and accurately, has shown modest improvements in reducing measurement noise but limited success in reducing placebo response per se [43].

10.6.3 | Extended Run-In Observation With Prospective Symptom Characterization

A particularly promising design-level strategy is the extended pre-treatment observation period. Applicable primarily to diseases that are not acutely progressing, this approach establishes a long run-in period before treatment initiation, during which each patient's principal symptoms and trajectories are carefully characterized. Investigators prospectively define measurement methods and clinically meaningful change criteria—whether stability, improvement from baseline, or slowing of functional loss. The observed individual trajectory is compared, where available, against natural history data for the same disease, strengthening interpretation of subsequent post-treatment changes. Critically, because the run-in period precedes any treatment, the pre-treatment trajectory is free of drug-related placebo contamination. Post-treatment changes can then be evaluated against this individually calibrated baseline: early improvements following the temporal signature of placebo (rapid onset, subsequent decay) are distinguishable from sustained pharmacological effects. This design directly addresses effort dependence concerns by establishing performance on effort-dependent measures during a period when no treatment-related motivational bias is present. One limitation of this approach deserves acknowledgment: the pre-treatment period is not entirely free of behavioral influences. Patients who have enrolled in a trial and are aware that treatment is imminent may underperform during run-in—either strategically or due to anxiety—or may exaggerate symptom severity, potentially inflating the apparent treatment effect when post-treatment assessments reflect their true capacity. Standardized assessment protocols, video-recorded evaluations, and objective physiological anchors (e.g., respiratory gas exchange during walk tests) help mitigate these effects, but cannot eliminate them entirely; this residual uncertainty should be acknowledged in trial reports.

10.7 | Blinded Outcome Assessment

Even in an otherwise open-label trial, outcome assessments can be conducted by evaluators blinded to the treatment timeline. This eliminates observer bias without requiring full double-blinding. For clinician-rated scales and imaging interpretation, blinded assessment is straightforward to implement. For example, video recordings of motor assessments can be rated by evaluators who do not know whether the recording was made before or after treatment initiation. This approach is already standard in several neuromuscular disease trials and should be considered routine in open-label rare disease studies.

10.7.1 | Integrating Multiple Methods

The methods described above are not mutually exclusive; their combined application provides converging evidence that is far more persuasive than any single method alone. A well-designed own-control trial might employ: (a) serial assessments enabling temporal trajectory modeling; (b) both objective and subjective endpoints for concordance validation; (c) Bayesian priors for formal placebo adjustment; (d) sensitivity analyses across a range of δ_p values; and (e) blinded outcome assessment for clinician-rated endpoints. When multiple independent analytical approaches converge on the same estimate of d_{drug}, the credibility of the placebo correction is substantially strengthened.

For objective endpoints, methods 1–3 (simple subtraction, temporal modeling, concordance) are typically sufficient, and the correction is minimal. For subjective endpoints, the full suite of methods 1–7 is available, and the analysis plan should specify which methods will be employed a priori. Table 3 summarizes the applicability of each method by endpoint type. Table 4 summarizes bias mechanisms in own-control designs and mitigation strategies.

11 | Discussion

The placebo effect has been cited as a fundamental objection to open-label and patient-as-own-control trial designs in rare diseases. The evidence and methods reviewed here allow a more precise assessment of this concern.

First, the placebo effect does not change the sample size required for either the RCT or the own-control design. Both are powered for d_{drug}, and the ratio between them (approximately 1:19 at $\rho = 0.90$, as demonstrated in table 3 of Summar and Woodcock [1]) is constant regardless of δ_p [1].

Second, the designs differ in how d_{drug} is confirmed after the trial. The RCT's concurrent control arm cancels placebo structurally, requiring no post-trial correction. The own-control design observes d_{total} and must subtract the estimated δ_p to isolate d_{drug}. This post-trial correction is the genuine additional step required by the own-control design.

Third, this correction is well-supported for objective endpoints. Placebo effects on enzyme activity, biomarker concentrations, molecular expression, imaging, hematologic parameters, and survival are consistently small and in most meta-analyses statistically indistinguishable from zero (SMD < 0.10). For the

17528062, 2026, 5, Downloaded from https://ascpt.onlinelibrary.wiley.com/doi/10.1111/cts.70550 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 | Bias mechanisms in own-control designs and mitigation strategies.

Bias mechanism	Description	Key design levers	Analytical mitigation
Placebo effect (expectation-driven)	Genuine physiological or psychological improvement from anticipation and therapeutic context; peaks early and decays	Serial assessments; blinded outcome assessment; extended run-in baseline	Temporal trajectory modeling; pre-specified sensitivity analysis with δ_p subtraction; Bayesian informative priors
Regression to the mean	Enrollment at symptomatic peak; subsequent measurements trend toward population mean regardless of treatment	Extended run-in to establish stable baseline; multiple pre-treatment assessments	Use of trajectory slope (not single baseline value) as comparator; natural history calibration
Hawthorne effect	Behavioral change from awareness of observation; more frequent monitoring may itself alter outcomes	Pre-treatment observation period matching post-treatment assessment frequency	Objective endpoint anchoring; concordance analysis between objective and subjective measures
Effort dependence	Increased patient motivation on volitional performance measures (6MWT, FVC, NSAA) after known treatment initiation	Extended run-in baseline on effort-dependent measures; standardized encouragement scripts; blinded assessors	Objective physiological anchors (gas exchange); temporal trajectory modeling to distinguish transient motivational gains from sustained improvement
Expectation bias (patient)	Optimistic self-reporting of subjective symptoms when treatment assignment is known	Objective primary endpoints; PRO anchored to objective measures; blinded run-in	Concordance analysis; growth mixture modeling to identify placebo-only responders
Observer/assessor bias	Clinician or caregiver ratings unconsciously influenced toward improvement when treatment is known	Blinded outcome assessment (evaluator unaware of treatment timeline); video-recorded assessments rated centrally	Pre-specified rater training; concordance with objective biomarker endpoints
Natural disease fluctuation	Waxing/waning disease course; improvement during treatment period may reflect natural cycle	Extended run-in to characterize individual trajectory; disease selection (stable or progressive preferred)	Natural history calibration; LMM with individual random slopes; sensitivity analyses under MNAR missingness

Note: Blinded outcome assessment and extended run-in baselines are highlighted as key design levers applicable across multiple bias mechanisms.
Abbreviations: 6MWT, 6-minute walk test; FVC, forced vital capacity; LMM, linear mixed model; MNAR, missing not at random; NSAA, North Star Ambulatory Assessment; PRO, patient-reported outcome.

majority of approved rare disease therapies, which use objective primary endpoints, the correction is minimal.

Fourth, for subjective endpoints where placebo effects are larger (SMD 0.20–0.50), the post-trial correction requires more care. However, the analytical toolkit available is far richer than simple subtraction. Temporal trajectory modeling exploits the decay signature of placebo effects. Concordance between objective and subjective measures provides internal validation. Bayesian methods formally incorporate uncertainty about δ_p. Growth mixture modeling and latent class analysis identify subpopulations of drug-only and placebo-only responders without requiring the additivity assumption. Machine learning–based prediction generates individualized placebo estimates. Causal inference methods provide principled adjustment when external controls are available. The combined application of these methods provides a level of analytic rigor that substantially addresses the interpretive gap between the two designs. Consultation with an experienced biostatistics team familiar with using these models is always recommended.

Fifth, effort dependence deserves specific attention. The FDA has cited effort dependence as a concern in reviewing rare disease applications, particularly for neuromuscular and pulmonary endpoints where patient exertion directly influences the measurement. In the own-control design, the extended run-in observation period establishes performance on effort-dependent measures before treatment-related motivational bias can operate. Post-treatment improvement can be evaluated against this individually calibrated baseline, and temporal trajectory analysis can distinguish sustained functional gains from transient motivational effects. In the small-sample RCT, by contrast, effort dependence is confounded with treatment-arm assignment whenever functional unblinding occurs—a common problem in rare disease trials where side effects or administration routes may differ perceptibly between arms.

17528062, 2026, 5, Downloaded from https://ascpt.onlinelibrary.wiley.com/doi/10.1111/cts.70550 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

Sixth, it bears emphasis that the RCT's structural advantage in handling placebo is itself compromised in small samples. As described in the preceding section, small-sample RCTs are vulnerable to asymmetric allocation of placebo responders across arms, an imbalance that is difficult to detect and correct without additional analyses. When combined with the recruitment impossibility documented elsewhere—where more than 95% of rare diseases cannot supply the patient numbers required for an adequately powered RCT—the practical choice is not between a perfect RCT and an imperfect own-control design, but between an underpowered RCT with hidden placebo vulnerabilities and a well-designed own-control study with explicit, pre-specified placebo management. For the vast majority of rare diseases, the latter is the more scientifically rigorous option.

Seventh, it is worth distinguishing two conceptually distinct problems that arise when an RCT is underpowered for rare disease. The first is a Type II error problem: the trial lacks sufficient statistical power to detect a true treatment effect, so real benefits go undetected and the trial is declared negative. The second is a bias problem: small samples allow asymmetric allocation of placebo responders to inflate or deflate the between-arm difference, distorting the point estimate of the drug effect even when a significant result is obtained. These two problems are separate in mechanism but compound each other in practice. An underpowered RCT not only risks missing a true effect (Type II error) but also produces a biased and imprecise estimate of the effect magnitude that is not correctable post hoc. The own-control design, by addressing both problems simultaneously—through patient-level efficiency that enables adequate power at small N and through pre-specified analytical adjustment for the placebo contribution—offers a more complete methodological solution for rare disease settings.

We acknowledge limitations. The placebo effect estimates in Table 1 derive primarily from common diseases; disease-specific estimates for individual rare diseases are rarely available. The additivity assumption, while standard, may not hold in all settings. Some of the advanced methods (GMM, ML prediction) require sample sizes or historical data that may not be available for ultra-rare diseases. Sensitivity analyses using a range of δ_p values should be standard practice in all cases.

Additionally, in conditions with highly variable disease courses, separating placebo from natural fluctuation requires robust natural history data—a challenge that underscores the importance of well-characterized disease registries [1]. The establishment of well-characterized natural history datasets, through initiatives such as the National Organization for Rare Disorders (NORD) IAMRARE registry platform and National Center for Advancing Translational Sciences (NCATS)–supported registries, can provide the disease-trajectory data needed to calibrate both natural fluctuation and placebo estimates.

12 | Conclusions and Recommendations

The placebo effect is a real but quantifiable phenomenon. Its magnitude varies predictably by endpoint type, reporter modality, and trial duration. It does not affect the power calculation or sample size for either design; rather, it determines the complexity of the post-trial analysis needed to confirm d_{drug} in the own-control design. We offer the following recommendations:

1. Sponsors should pre-specify placebo effect estimates and the analytical methods for placebo adjustment in their statistical analysis plans, derived from meta-analytic data and matched to the endpoint type and reporter modality used in their trial.
2. For objective endpoints, simple sensitivity analysis with placebo subtraction is sufficient. The correction is small ($\delta_p < 0.10$) and any imprecision is inconsequential.
3. For subjective endpoints, sponsors should employ multiple complementary methods—temporal trajectory modeling, concordance with objective measures, Bayesian priors, and where feasible, growth mixture modeling or latent class analysis—to provide converging evidence for the drug effect estimate.
4. Trials should incorporate serial assessments to capture temporal dynamics of improvement, enabling distinction between sustained pharmacological effects and decaying placebo responses.
5. Objective endpoints should be included alongside subjective measures whenever possible, providing internal concordance validation.
6. Blinded outcome assessment should be implemented for clinician-rated endpoints, even in otherwise open-label designs.
7. Regulatory agencies should develop explicit guidance on acceptable methods for placebo adjustment in open-label rare disease trials, including the advanced methods described here, providing sponsors with a clear framework. The FDA's existing draft guidance on Bayesian methodology in clinical trials already endorses the use of informative priors incorporating external data [32], and several of the causal inference and trajectory modeling approaches described in this article are consistent with FDA's stated willingness to consider complex innovative designs in the rare disease setting. While the details of individual NDA submissions are not publicly documented, sponsors should engage these methods prospectively with health authorities through Type B or Type C meetings, where alignment on acceptable analytical approaches can be established before trial conduct. The CONSORT 2025 guidelines now emphasize detailed reporting of placebo characteristics and harms assessment in all trial reports, establishing a precedent for greater transparency in placebo handling [44].
8. Sensitivity analyses exploring a range of placebo estimates should be standard, with results reported transparently to demonstrate the robustness of treatment effect conclusions.
9. For trials employing effort-dependent endpoints, an extended pre-treatment run-in observation period (as described in Method 6) should be incorporated to establish each patient's baseline performance in the absence of

17528062, 2026, 5, Downloaded from https://ascpt.onlinelibrary.wiley.com/doi/10.1111/cts.70550 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

treatment-related motivational bias, directly addressing a concern the FDA has cited in neuromuscular and pulmonary rare disease applications.

10. Regulatory reviewers and sponsors should recognize that in rare disease settings, the RCT's structural advantage in canceling placebo effects is itself degraded by small sample sizes, where asymmetric allocation of placebo responders can distort between-arm comparisons in ways that are difficult to detect and correct without additional analyses. An underpowered RCT is not a safer or more rigorous choice than a properly designed own-control trial; it is a less informative one.

The choice of trial design in rare disease is constrained by patient availability. The RCT handles placebo automatically—a genuine structural advantage—but at an approximately 19-fold cost in patients (at $\rho = 0.90$) [1]. The patient-as-own-control design requires a post-trial placebo correction—an additional analytic step—but gains an enormous efficiency advantage through within-subject correlation. The analytical methods available for this correction are not ad hoc; they represent decades of rigorous development in statistical methodology and placebo science. For the majority of rare diseases, the own-control design with a rigorous post-trial correction represents the optimal path forward, because the RCT's patient requirement exceeds the available population.

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. M. L. Summar and J. Woodcock, "Supply and Demand in the Mathematics of Rare Disease Drug Development: Why Choosing the Right Model Is Crucial," *Clinical and Translational Science* (2026).

2. 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.

3. 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.

4. 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.

5. A. Hróbjartsson and P. C. Gøtzsche, "Placebo Interventions for All Clinical Conditions," *Cochrane Database of Systematic Reviews* 1 (2010): CD003974.

6. T. D. Wager and L. Y. Atlas, "The Neuroscience of Placebo Effects: Connecting Context, Learning and Health," *Nature Reviews. Neuroscience* 16, no. 7 (2015): 403–418.

7. A. G. Barnett, J. C. van der Pols, and A. J. Dobson, "Regression to the Mean: What It Is and How to Deal With It," *International Journal of Epidemiology* 34, no. 1 (2005): 215–220.

8. J. McCambridge, J. Witton, and D. R. Elbourne, "Systematic Review of the Hawthorne Effect: New Concepts Are Needed to Study Research Participation Effects," *Journal of Clinical Epidemiology* 67, no. 3 (2014): 267–277.

9. B. Colagiuri, L. A. Schenk, M. D. Kessler, S. G. Dorsey, and L. Colloca, "The Placebo Effect: From Concepts to Genes," *Neuroscience* 307 (2015): 171–190.

10. A. Hróbjartsson, A. S. Thomsen, F. Emanuelsson, et al., "Observer Bias in Randomized Clinical Trials With Measurement Scale Outcomes: A Systematic Review of Trials With Both Blinded and Nonblinded Assessors," *Canadian Medical Association Journal* 185, no. 4 (2013): E201–E211.

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

12. J. C. Fendel, C. Tiersch, P. Sölder, J. Gaab, and S. Schmidt, "Effects of Open-Label Placebos Across Populations and Outcomes: An Updated Systematic Review and Meta-Analysis of Randomized Controlled Trials," *Scientific Reports* 15 (2025): 29940, https://doi.org/10.1038/s41598-025-14895-z.

13. T. A. Mestre, M. Shamy, F. Benedetti, and A. E. Lang, "Harnessing the Power of Placebos in Movement Disorders: Insights From Parkinson's Disease in Clinical Research and Practice," *Movement Disorders* 33, no. 8 (2018): 1195–1203, https://doi.org/10.1002/mds.27409.

14. P. J. Brooks, D. A. Tagle, and S. Groft, "Expanding Rare Disease Drug Trials Based on Shared Molecular Etiology," *Nature Biotechnology* 32, no. 6 (2014): 515–518.

15. T. Bschor, H. Kern, J. Henssler, and C. Baethge, "Differential Outcomes of Placebo Treatment Across 9 Psychiatric Disorders: A Systematic Review and Meta-Analysis," *JAMA Psychiatry* 81, no. 7 (2024): 700–709.

16. R. Botvinik-Nezer, B. Petre, M. Ceko, M. A. Lindquist, N. P. Friedman, and T. D. Wager, "Placebo Treatment Affects Brain Systems Related to Affective and Cognitive Processes, but Not Nociceptive Pain," *Nature Communications* 15 (2024): 6017, https://doi.org/10.1038/s41467-024-50103-8.

17. M. Von Wernsdorff, M. Loef, B. Tuschen-Caffier, and S. Schmidt, "Effects of Open-Label Placebos in Clinical Trials: A Systematic Review and Meta-Analysis," *Scientific Reports* 11 (2021): 3855.

18. J. S. M Lidstone, E. Meins, C. Fernyhough, et al., "The roles of private speech and inner speech in planning during middle childhood: Evidence from a dual-task paradigm," *Journal of Experimental Child Psychology* 107, no. 4 (2010): 438–451, https://doi.org/10.1016/j.jecp.2010.06.002. PMID: 20633894.

19. M. Ahmadzad-Asl, J. Javan-Noughabi, L. Z. Atkinson, et al., "Placebo Response in Antidepressant Trials: A Systematic Review and Meta-Analysis," *Journal of Affective Disorders* 306 (2022): 237–247.

20. L. Vase, G. L. Petersen, J. L. Riley, 3rd, and D. D. Price, "Factors Contributing to Large Analgesic Effects in Placebo Mechanism Studies Conducted Between 2002 and 2007," *Pain* 145, no. 1–2 (2009): 36–44.

21. X. Chen, K. Zou, N. Abdullah, et al., "The Placebo Effect and Its Determinants in Fibromyalgia: Meta-Analysis of Randomised Controlled

17528062, 2026, 5, Downloaded from https://ascpt.onlinelibrary.wiley.com/doi/10.1111/cts.70550 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

Trials," *Clinical Rheumatology* 36, no. 7 (2017): 1623–1630, https://doi.org/10.1007/s10067-017-3595-8.

22. 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.

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

24. C. Eiser and J. W. Varni, "Health-Related Quality of Life and Symptom Reporting: Similarities and Differences Between Children and Their Parents," *European Journal of Pediatrics* 170, no. 10 (2011): 1203–1208.

25. P. Upton, J. Lawford, and C. Eiser, "Parent-Child Agreement Across Child Health-Related Quality of Life Instruments: A Review of the Literature," *Quality of Life Research* 17, no. 6 (2008): 895–913.

26. R. Meister, M. Abbas, J. Antel, et al., "Placebo Response Rates and Potential Modifiers in Double-Blind Randomized Controlled Trials of Second and Newer Generation Antidepressants for Major Depressive Disorder," *European Neuropsychopharmacology* 30 (2020): 1–13.

27. 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.

28. P. Enck, U. Bingel, M. Schedlowski, and W. Rief, "The Placebo Response in Medicine: Minimize, Maximize or Personalize?," *Nature Reviews. Drug Discovery* 12, no. 3 (2013): 191–204.

29. M. Curković, A. Košec, and A. Savić, "Re-Evaluation of Significance and the Implications of the Placebo Effect in Antidepressant Therapy," *Frontiers in Psychiatry* 10 (2019): 143.

30. D. J. Spiegelhalter, K. R. Abrams, and J. P. Myles, *Bayesian Approaches to Clinical Trials and Health-Care Evaluation* (Wiley, 2004).

31. 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.

32. U.S. Food and Drug Administration, *Use of Bayesian Methodology in Clinical Trials for Drug and Biological Products: Draft Guidance* (U.S. Food and Drug Administration, 2026).

33. B. Muthén and H. C. Brown, "Estimating Drug Effects in the Presence of Placebo Response: Causal Inference Using Growth Mixture Modeling," *Statistics in Medicine* 28, no. 27 (2009): 3363–3385.

34. R. Kessels, R. Mozer, and D. B. Rubin, "Methods for De-Identifying Placebo and Drug Group Longitudinal Data Within the Context of a Clinical Trial," *Statistics in Medicine* 38, no. 9 (2019): 1581–1598.

35. T. Tarpey, E. Petkova, and R. T. Ogden, "Profiling Placebo Responders by Self-Consistent Partitions of Clinical Trials," *Statistics in Medicine* 22, no. 20 (2003): 3373–3390.

36. Z. Bakk and J. Kuha, "Two-Step Estimation of Models Between Latent Classes and External Variables," *Psychometrika* 83 (2018): 871–892.

37. Cognivia SA, "Placebell Method: Improving Clinical Trial Assay Sensitivity Through Machine Learning–Based Placebo Prediction. White Paper," (2023), https://www.cognivia.com.

38. P. R. Rosenbaum and D. B. Rubin, "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika* 70, no. 1 (1983): 41–55.

39. J. M. Robins, A. Rotnitzky, and L. P. Zhao, "Estimation of Regression Coefficients When Some Regressors Are Not Always Observed," *Journal of the American Statistical Association* 89, no. 427 (1994): 846–866.

40. J. K. Lunceford and M. Davidian, "Stratification and Weighting via the Propensity Score in Estimation of Causal Treatment Effects: A Comparative Study," *Statistics in Medicine* 23, no. 19 (2004): 2937–2960.

41. M. J. van der Laan and S. Rose, *Targeted Learning: Causal Inference for Observational and Experimental Data* (Springer, 2011).

42. J. M. Snowden, S. Rose, and K. M. Mortimer, "Implementation of G-Computation on a Simulated Dataset: Demonstration of a Causal Inference Technique," *American Journal of Epidemiology* 173, no. 7 (2011): 731–738.

43. S. Raman, "Statistical Methods in Handling Placebo Effect," in *Neuroscience of Placebo Effects*, ed. L. Colloca (Academic Press, 2020), 167–185.

44. N. J. Butcher, A. Monsour, E. Gargon, et al., "CONSORT 2025 Explanation and Elaboration: Updated Guideline for Reporting Randomised Trials," *BMJ* 389 (2025): e082914.

17528062, 2026, 5, Downloaded from https://ascpt.onlinelibrary.wiley.com/doi/10.1111/cts.70550 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