The 100-to-1 Imbalance: Recovering Statistical Power in Rare Disease Trials Through Global Statistical Tests (GSTs)

Samuel Dickson PhD, Josh Christensen MS, Daisy Harris MS, Newman Knowlton MS, Suzanne Hendrix PhD; Pentara Corporation, Salt Lake City, Utah, USA

Abstract IR020

We are currently 178 times more likely to reject a treatment that works than to approve a treatment that doesn't work.

BETA/ALPHA IMBALANCE

Standard trial design (α=2.5%, β=20%) has a β/α ratio of 8. So, **approving an ineffective treatment is considered 8 times worse than abandoning a treatment that works.**

The Rare Disease Penalty

TYPE 1 ERROR (FALSE POSITIVE) ≤ 2.5%

Risk of approving an ineffective therapy

- Secondary endpoint consistency required
- Effect must be clinically meaningful
- Subgroups must show consistent effect
- Sensitivity analyses must be consistent even with worst case imputation

α Heavily Protected (often < 0.1%)

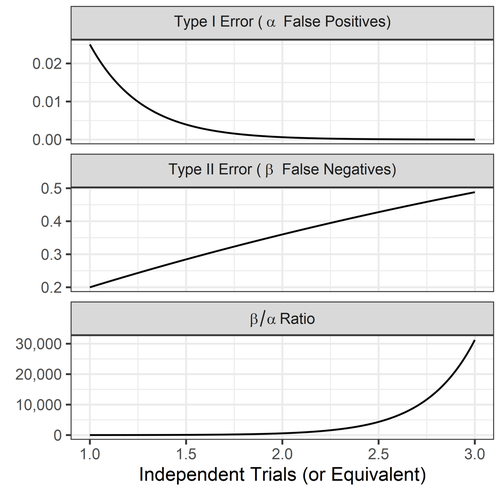
TYPE 2 ERROR (FALSE NEGATIVE) > 20%

Risk of abandoning a truly working cure

- Wrong primary outcome selected
- Scale is too discrete ("Clinically Meaningful")
- Can't get enough patients or follow-up time
- Primary outcome doesn't progress enough
- Disease is heterogeneous

β Inflation Routinely Tolerated (often >35%)

Empirical Evidence:
A Pentara review of treatments for rare diseases that received CRLs shows the median estimated ratio of Beta/Alpha is 178 for unapproved therapies, meaning an ineffective drug approval is treated as more than 100× worse than disease progression.



Trials (or equivalent)	Type I Error	Type II Error	β/α Ratio
1	0.025	0.200	8
2	0.000625	0.360	576
3	0.0000156	0.488	31,232

Who bears the cost? Patients

A patient-centered framework prioritizes patient benefit over fear of approving a treatment that may not work.

Response Percentile	Two-Sided P-Value	Clinical Meaning	
99.5%	p = 0.01	Peak Efficacy	A+
97.5%	p = 0.05	Standard Threshold (α)	A+
95%	p = 0.10	Trend Threshold (Failed)	A
90%	p = 0.20	Positive Shift (Failed)	A-
85%	p = 0.30	Moderate Benefit	B
75%	p = 0.50	Moderate Benefit	C
65%	p = 0.70	Small Benefit	D
50%	p = 1.00	Null (Zero Effect)	F
35%	p = 0.70	Small Decline	F
25%	p = 0.50	Moderate Decline	F
15%	p = 0.30	Moderate Decline	F
10%	p = 0.20	Negative Shift	F
5%	p = 0.10	Trend Worsening	F
2.5%	p = 0.05	Significant Worsening	F
0.5%	p = 0.01	Severe Harm	F

	Treatment Works	Treatment Does Not Work
Trial Succeeds	True Positive	Type 1 Error (False Positive) Regulator/Payer penalty
Trial Fails	Type 2 Error (False Negative) Patient penalty	True Negative

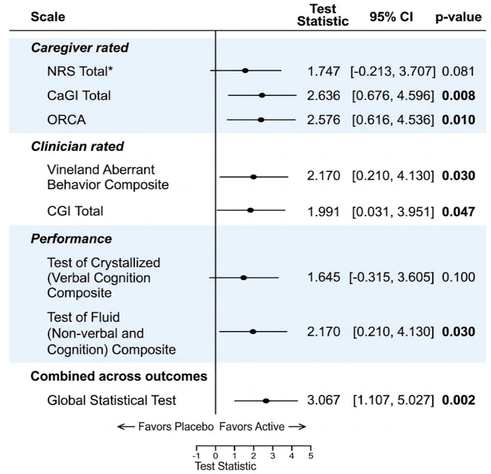
- **90th Percentile Trap:** A strong 90th percentile response yields p = 0.20 and is discarded as trial failure.
- **Two-sided Equivalence:** P-values must be low with an effect in the wrong direction to indicate harm, but that symmetry is not recognized.

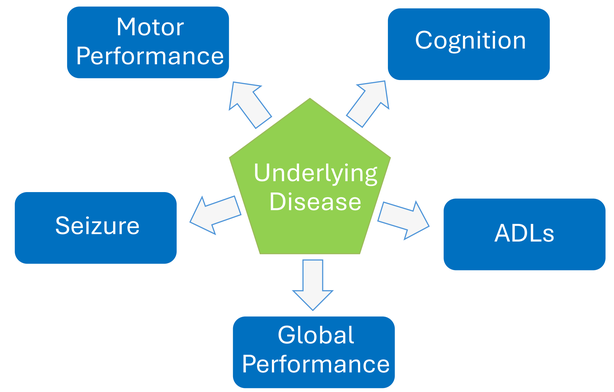
While statistical rigor protects against false approvals, it does not acknowledge the tradeoff, that the price is paid by patients when viable therapies are abandoned.

GLOBAL STATISTICAL TESTS

GSTs allow us to recover statistical power by capturing multi-domain efficacy.

- **Solves the Small-N Power Trap:** Recovers statistical power by aggregating therapeutic signals across domains without inflating Type 1 error.
- **Eliminates Endpoint Guesswork:** Removes the risk of betting a trial on a single primary endpoint when lack of natural history data makes progression unpredictable.
- **Reflects Multi-System Biology:** Measures total disease modification rather than isolated symptoms, capturing concurrent impacts on cognition, motor, behavior, function and seizures.
- **Rescues "Falsely Failed" Trials:** Empirical evidence shows rare disease programs often appear to have ambiguous results but actually demonstrate unified statistical efficacy overall.





REGULATORY FLEXIBILITY IN PRACTICE

1. Prespecified Global Primary Endpoints
Accept GSTs and composites as standalone primary endpoints.

2. One-Sided Alpha Testing
Adopt one-sided testing frameworks (α=0.05) and show percentiles to align with therapeutic benefit.

3. Use a Time-Savings Metric
Adopt "Time Saved" (delay in disease progression) as the benchmark for clinical meaningfulness.

4. Heterogeneity Tolerance
Evaluate total disease benefit without requiring universal consistency across sub-domains, subgroups, and sensitivity analyses.

5. Compare Benefit:Risk Ratio to Beta:Alpha Ratio: Evaluate ratios across the entire development program, and align approval decision with Benefit:Risk assessment.

REFERENCES

1. O'Brien, P. C. (1984). Procedures for comparing samples with multiple endpoints. *Biometrics*, 40(4), 1079–1087.
2. U.S. Food and Drug Administration. (2022). *Multiple Endpoints in Clinical Trials (Guidance for Industry)*.
3. U.S. Food and Drug Administration. (2023). *Rare Diseases: Considerations for the Development of Drugs and Biological Products (Guidance for Industry)*.

CONTACT

Samuel P Dickson, PhD
sdickson@pentara.com

Pentara