

Supplementary Methods and Simulation Results

Companion to Testing Shared and Differentiated Functional Architectures Across Psychiatric Conditions. All computational results below are synthetic; no participant data were analyzed.

S1. Estimands and architectural models

In the primary target population, B contains standardized diagnostic case-control effects. Each column uses the same age distribution and covariate definition. At site s , deviations from B include both site-specific departures and sampling error. Missing contrasts remain missing. Shared-control covariance belongs to the sampling model; individual repeat scans are not independent participants.

$$\hat{B}_s = B + U_s + \varepsilon_s$$

$$\text{Shared: } B = g a^T + R$$

$$\text{Shared + differentiated: } B = g a^T + H C^T + R$$

$$\text{Differentiated: } B = H C^T + R$$

Normalize the common edge direction g to unit root-mean-square loading when amplitudes are interpreted on the standardized effect scale. Conventional matrix normalization to unit Euclidean length is equivalent after rescaling a . Constrain fitted amplitudes to be nonnegative only for the common-direction candidate. Signed validation projections remain unconstrained. Decompositions may use orthogonality for identification, but a singular vector or positive diagnostic average does not identify a unique common cause.

For disorder d , $L(d, M)$ is the squared error averaged across its held-out site contrasts and E features. Aggregate all outer-fold predictions for that disorder before division. $L(d, 0)$ is the corresponding mean squared observed effect. The primary macro score is the mean of the D disorder-specific predictive R^2 values. Equal weights apply to disorders, then to eligible sites within each disorder. This is not the ratio of mean losses and is not participant-pooled R^2 .

$$R^2_{macro, M} = \frac{1}{D} \sum_{d=1}^D \left[1 - \frac{L(d, M)}{L(d, 0)} \right]$$

For the simulation, estimate stable signal $S(d)$ by averaging the feature-wise inner products of observed effect maps from every pair of distinct sites contributing disorder d . Independent zero-mean site and sampling deviations then vanish in expectation. This assumption fails for correlated acquisition bias across sites, which is one reason the nuisance benchmark remains necessary. The ceiling is the average of $S(d)/L(d, 0)$, using the primary diagnostic weights. The signal numerator estimates recoverable population effects rather than within-site repeatability. Both numerator and denominator are random; ceiling uncertainty must be retained.

$$C_{macro} = \frac{1}{D} \sum_{d=1}^D \frac{S(d)}{L(d, 0)}$$

The reported simulation estimator uses all eligible sites for evaluation-only cross-products. Model directions, ranks, and shrinkage are already frozen. A real analysis may use independent sites to estimate the ceiling. A within-site split-half estimator alone counts site-specific stable differences as reproducible and therefore targets another ceiling. Site bootstrapping estimates uncertainty under the assumed site sampling process, not immunity to systematic bias. With only three sites, diagnostic ceilings can be highly unstable.

S2. Exact illustrative decision rules and implementation

The simulation uses a mean signal-energy floor $v = 0.0025$, a fraction-of-ceiling floor $p = 0.20$, a non-inferiority multiplier $\kappa = 0.10$, and a minimum signed projection amplitude of 0.02. These values were chosen to illustrate declared resolution, not derived as optimal or clinically validated constants. The generating configurations and decision settings were fixed for the reported 200-repetition run after code checks. No formal preregistration was performed.

For each repetition, generate 399 bootstrap site-weight draws from a multinomial distribution with 12 trials and equal site probabilities. All contrasts at a resampled site receive the same weight, preserving shared-control dependence. Exclude draws without a diagnostic group and exclude draws with fewer than two distinct sites for a diagnostic cross-product. Retain the resulting valid draws rather than pretending that missing coverage is informative. Model predictions are not refitted in these bootstrap draws. Bounds are conditional, exploratory, and not claimed to attain nominal coverage.

A detectable result requires lower 5th-percentile bounds above zero for the estimated ceiling, signal energy minus v , and the better candidate score minus p times the ceiling. The better score is the maximum of the two already cross-validated candidate scores in each resampled evaluation. Because this maximum and the conditional resampling can be optimistic, the simulation does not establish error control for this gate. A real confirmatory analysis should calibrate the complete selection-aware rule or test candidate floors with appropriate multiplicity handling.

Support no reproducible structure above the chosen resolution when the upper 95th-percentile bound for signal energy is below v , or the upper bound for the better-candidate-minus-fractional-ceiling quantity is below zero. If neither this result nor the positive evidence gate is supported, report unresolved. For rank-one sufficiency, require its own floor to pass and the upper 95th-percentile bound for higher-rank gain minus κ times the ceiling to be negative. For a material gain, require the lower 5th-percentile bound for that same contrast to be positive. This is one-sided non-inferiority, not symmetric equivalence.

For independent broad transfer, group the eight synthetic diagnoses into four pairs before fitting. In each outer split, learn a normalized direction from the training-site mean maps of the other three pairs. Project each held-out diagnostic effect map onto that direction. Construct t-based one-sided bounds across contributing sites with Bonferroni adjustment over eight groups, using degrees of freedom equal to site count minus one. A broad pass requires all lower bounds above 0.02. A precise failure requires at least one adjusted upper bound below 0.02. Shared: detectable, rank one sufficient, broad pass. Shared plus differentiated: detectable, material gain, broad pass. Differentiated: detectable, material gain, precise failure. Other cases remain unresolved. With few sites and overlapping training sets, these t bounds are approximate and their coverage is not validated.

This broad-transfer test is stronger than the basic same-disorder loading test. It depends on independently declared families. The experiment uses generating-family pairs to make the distinction transparent, a favorable assumption. It does not test all possible alternative family groupings. The reported categorical implementation uses rank comparison and family transfer; the frozen-component ablation discussed in the perspective is a complementary analysis, not an additional validated gate in this experiment.

There are 63 observed site-by-disorder contrasts. Site coverage by diagnostic group is 12, 12, 9, 9, 6, 6, 6, and 3. Rare-group sites are 1, 5, and 9 in one-based numbering. Outer folds are consecutive site blocks 1–4, 5–8, and 9–12. Each inner training partition therefore retains one rare-group site, and each outer training partition retains two. Generalized least-squares estimates use known generating covariance, including reused controls and site-shared variance. Low-rank compression then treats diagnostic columns equally, with inner folds selecting shrinkage and rank. This simplified estimator is not the full covariance-weighted low-rank optimizer proposed for a consortium analysis.

Edge covariance has eight blocks of eight features, within-block correlation 0.25, and unit marginal variance. Site-shared Gaussian deviations have standard deviation 0.04; contrast-specific deviations have standard deviation 0.02. Gaussian case and control mean errors use variances $1/n$. At four sites, the same control mean is subtracted from

every diagnostic case mean. Generated sample sizes and every matrix are stored in `design.npz`. The common direction and seven orthogonal feature patterns are drawn once by seeded QR decomposition; amplitudes and differential components are explicit in `simulate.py`.

The nuisance-only scenario adds an independently drawn orthogonal template with amplitude $0.20 + 0.08$ times a standard normal draw at each observed contrast. This term is shared in expectation and can transfer across sites despite zero stable diagnostic effects. Oracle adjustment subtracts exactly the generating nuisance term. This demonstrates an identifiability hazard; it is not evidence that real motion or vigilance can be removed perfectly. A practical extension must estimate nuisance maps only in independent reference or training data and propagate their measurement error.

S3. Repetition-level variability and outcomes

Table S1. Score variability across 200 independent repetitions

Scenario	Estimated macro ceiling	Rank-one R^2	Higher-rank R^2
No stable effect	-0.000 [-0.030, 0.036]	-0.015 [-0.033, 0.003]	-0.027 [-0.049, -0.004]
Shared	0.675 [0.650, 0.694]	0.648 [0.618, 0.669]	0.621 [0.599, 0.641]
Common + differentiated	0.765 [0.747, 0.783]	0.464 [0.438, 0.491]	0.720 [0.692, 0.751]
Differentiated families	0.738 [0.719, 0.762]	0.138 [0.107, 0.177]	0.678 [0.656, 0.712]
Weak shared	0.031 [-0.000, 0.064]	-0.011 [-0.030, 0.007]	-0.002 [-0.025, 0.025]
Nuisance only	0.620 [0.552, 0.671]	0.569 [0.503, 0.632]	0.548 [0.461, 0.611]
Nuisance, oracle-adjusted	-0.001 [-0.022, 0.030]	-0.016 [-0.031, 0.005]	-0.027 [-0.048, -0.004]

Brackets are empirical 2.5th–97.5th percentiles across independently generated repetitions, not confidence intervals for one dataset. Scores and ceilings share the same diagnostic weighting. Conditional within-repetition bootstrap bounds are distinct from these Monte Carlo distributions.

Table S2. Exploratory operational classifications, counts out of 200

Scenario	No structure	Shared	Shared + differentiated	Differentiated	Unresolved
No stable effect	190	0	0	0	10
Shared	0	180	0	0	20
Common + differentiated	0	0	138	0	62
Differentiated families	0	0	0	196	4
Weak shared	180	0	0	0	20
Nuisance only	0	21	0	0	179
Nuisance, oracle-adjusted	199	0	0	0	1

No structure is qualified by the selected signal floor. For a rate p based on 200 repetitions, Monte Carlo standard error is $\sqrt{p(1-p)/200}$, at most 0.0354. Zero observed errors do not establish a zero error probability. These rates characterize this simplified implementation under its stated assumptions.

Table S3. Estimated ceilings by diagnostic group in two signal scenarios

Group	Sites	Shared ceiling	Monte Carlo range	Weak ceiling	Monte Carlo range
D1	12	0.594	[0.548, 0.633]	0.019	[-0.021, 0.057]
D2	12	0.621	[0.580, 0.660]	0.024	[-0.014, 0.070]
D3	9	0.671	[0.623, 0.711]	0.029	[-0.021, 0.096]

Group	Sites	Shared ceiling	Monte Carlo range	Weak ceiling	Monte Carlo range
D4	9	0.697	[0.654, 0.742]	0.032	[-0.018, 0.086]
D5	6	0.743	[0.701, 0.784]	0.039	[-0.037, 0.120]
D6	6	0.765	[0.728, 0.805]	0.046	[-0.039, 0.132]
D7	6	0.809	[0.774, 0.840]	0.049	[-0.027, 0.127]
D8	3	0.499	[0.368, 0.602]	0.007	[-0.158, 0.188]

Ranges are repetition-to-repetition empirical 2.5th–97.5th percentiles. Negative estimates are retained. This table reports ceilings, not individual clinical prediction accuracy. Full disorder-level ceiling summaries for the other generating scenarios are in `ceilings_by_disorder.csv`.

S4. Verification of factual and bibliographic details

Initiative announcement. Sidney Taiko Sheehan; October 22, 2026. Public USC release explicitly contains “targeted multiverse approach.” <https://keck.usc.edu/news/global-team-launches-largest-ever-study-of-brain-function-across-eight-mental-illnesses/>

NIH award. NDA public grant fields list R01MH142539-01, August 7, 2026–July 31, 2031. Its outbound public RePORTER project is 11269134. The main reference uses RePORTER. <https://reporter.nih.gov/project-details/11269134>

Sha et al. (2019). Resting-state meta-analysis: 8,298 patients and 8,165 controls. Structural sample sizes are different and not substituted. <https://pubmed.ncbi.nlm.nih.gov/30612699/>

Xia et al. (2018). Discovery 663; independent replication 336. <https://pubmed.ncbi.nlm.nih.gov/30068943/>

Kebets et al. (2019). 224 total participants; partial least squares. <https://pubmed.ncbi.nlm.nih.gov/31515054/>

Cocuzza et al. (2026). Nature Communications 17, 6678; July 21, 2026. Diagnosed n=134; without diagnosis n=85; total 219. <https://doi.org/10.1038/s41467-026-75585-6>

Rutherford et al. (2023). eLife 12:e85082 explicitly extends normative models to functional connectivity; Yeo-17 and Smith-10 representations. Retained as directly relevant. <https://elifesciences.org/articles/85082>

McTeague et al.. 2017 AJP paper: cognitive control; 2020 AJP paper: emotional processing. Both are task-activation precedents. <https://doi.org/10.1176/appi.ajp.2017.16040400>

S5. Reproducibility materials and limits

The companion archive contains `simulate.py`, `summarize.py`, `results.csv`, `summary.json`, `ceilings_by_disorder.csv`, `design.npz`, `requirements.txt`, `README.md`, and `checks.json`. Run `simulate.py` with `SIM_REPS=200` and `OPENBLAS_NUM_THREADS=1`, then run `summarize.py`. The fixed seed is 20261023; a separate fixed bootstrap seed is 20261024. The code depends on NumPy and SciPy. Runtime and package versions are recorded in the archive. Figure 1 is schematic and Figure 2 is procedural; both are embedded in the main Word document and are not counted as simulation results.

The experiment does not model BOLD time series, preprocessing error, developmental trajectories, unmeasured confounding, empirical atlas topology, comorbidity, treatment selection, or clinical outcomes. It provides calculated evidence about rank and transfer discrimination under a transparent synthetic design. It cannot determine the architecture, effect sizes, or power of the announced initiative. Broader sensitivity and coverage studies remain required before confirmatory application.