International Journal of Social Sciences and English Literature
Vol. 10, No. 4, 6-12, 2026
ISSN(E) 2576-683X
DOI: 10.55220/2576-683x.v10.920

Signature Over Substance: Homeopathy as a Diagnostic-Moderated Information Intervention

Elias Rubenstein

Independent Researcher, Fort Lauderdale FL 33309, USA.

Abstract

Homeopathy remains scientifically contested because ultra-high dilutions challenge substance-based causal models and reported outcomes are heterogeneous across indications and study settings. We translate the controversy into preregisterable discrimination tests that separate non-specific context effects from any diagnosis-specific add-on attributable to a structured intervention signature under matched context. The framework is mechanism-open and defines a preregistered primary outcome as the change in a prespecified deviation proxy (a composite is allowed with weights fixed a priori). The key prediction is diagnostic moderation: if a diagnosis-specific add-on exists, it should strengthen as preregistered diagnostic certainty increases, where diagnostic certainty is computed independently of outcomes (preferred: a preregistered latent-class measurement model; fallback: prespecified agreement-based proxy strata). The proposed matched-context three-arm design compares a diagnosis-congruent protocol, an otherwise identical placebo, and a diagnosis-incongruent mismatch protocol; the mismatch arm is governed by an explicit equivalence margin as a falsification device for expectancy artifacts. Validation combines an in silico suite (type-I error, power, and a blinding-leakage stress test) with a secondary consistency check using published risk-of-bias reliability strata. This manuscript is a modelling and preregistration framework with simulation validation; it does not claim clinical efficacy.

Keywords: Diagnostic Moderation, Equivalence Margin, Homeopathy, In Silico Validation, Information Medicine (Operational), Latent-Class Measurement, Matched-Context Trial Design, Placebo/Context Effects, Randomized Controlled Trials.

1. Introduction

Homeopathy is debated for a simple reason: the evidence landscape is heterogeneous, and the dominant explanatory model in pharmacology is substance-centric. Some trials and clinical reports suggest small improvements beyond placebo in certain indications, while other analyses find no stable add-on effect once stricter quality criteria are applied (Linde et al., 1997; Shang et al., 2005; Mathie et al., 2014; Mathie et al., 2017). If an add-on effect exists, it is unlikely to be uniform across diagnoses, settings, or intervention choices. In addition, ultra-high dilutions raise a mechanistic plausibility objection within substance-based pharmacological models and are a central reason why homeopathy remains scientifically contested (EASAC 2017; Grams 2019).

This paper takes a practical route. Instead of arguing plausibility first, it asks: *What preregistered pattern of results would have to appear for a diagnosis-specific add-on component to be credible under strict matched context?* The contribution is therefore not a new clinical claim, but a testable hypothesis framework with explicit decision and falsification conditions.

Six concrete contributions (operational, preregistration-ready). (1) A mechanism-open, *operational* definition of "information medicine" as reproducible preregistered outcome differences under matched context, without assuming a physical carrier. (2) A primary preregistered endpoint $Y = \Delta D^*$ defined as a change in a preregistered deviation proxy (composite allowed, weights fixed a priori). (3) An outcome-based add-on effect definition $E_{\text{spec}}(S_{\text{diag}})$ that cleanly separates any diagnosis-specific add-on from non-specific context effects. (4) A preregistered, outcome-independent operationalization of diagnostic certainty S_{diag} (preferred: measurement-model correctness probability; fallback: prespecified agreement-based proxy strata). (5) A matched-context three-arm discriminative design (Qc vs P vs Qm) where the mismatch arm is governed by an explicit equivalence margin δ as a falsification device for context/expectancy artifacts. (6) A validation pathway combining an in silico stress-test suite (type-I error, power, leakage) with a secondary external consistency check using published sensitivity patterns (risk-of-bias reliability strata).

1.1. A Motivating Anecdote (Non-Evidential)

Practitioners report a familiar tension: two patients with superficially similar complaints can respond very differently to the same contextual care environment, and the alleged "specific" component is claimed to depend on how precisely the disturbance pattern is identified. This manuscript does not treat such reports as evidence. It uses them only as motivation to define a design in which that claim becomes experimentally decidable.

2. Evidence Landscape as the Starting Point

Across indications, placebo-controlled trials yield mixed outcomes and meta-analyses diverge with methodological choices (Linde et al., 1997; Shang et al., 2005; Mathie et al., 2014; Mathie et al., 2017). A systematic review of meta-analyses reports overall effects versus placebo across indications with attenuation under stricter quality filters (Hamre et al., 2023). This profile does not resolve the controversy; it motivates moderator-based discrimination: specify the condition under which any diagnosis-specific add-on should strengthen, and design a study that can falsify that condition.

In addition to randomized evidence, the homeopathy literature contains structured case documentation and programmatic "best case" style evaluations. The Banerji protocols were reviewed in the context of the U.S. National Cancer Institute (NCI) Best Case Series Program (Banerji et al., 2008), and subsequent discussions of homeopathy and NCI best-case standards have been published in integrative oncology venues (Olaku et al., 2013). These streams do not settle efficacy, but they motivate a modelling stance: translate qualitative or case-based claims about diagnosis-specificity into preregistered discriminative patterns that can fail clearly under strict controls.

2.1. Framework Postulate

The manuscript adopts a working postulate: biological organization can be treated in informational terms for the purpose of generating falsifiable predictions. Acceptance of the postulate is not required; only preregistered predictions and test outcomes matter. In this manuscript, "information" is strictly an *operational label* for reproducible, preregistered outcome differences under matched context; it does not presume any specific physical substrate.

3. Conceptual Model and Working Definitions

An organism is represented as a high-dimensional biological state distribution ρ_{bio}. Health corresponds to an individual reference attractor σ_Z^{bio}: a stable functional regime characterized by coherent biological patterns. Disease is treated as a persistent deviation from this regime, summarized by an incoherence or distance functional $D(\rho_{\text{bio}} \| \sigma_Z^{\text{bio}})$.

Therapy is modeled as an intervention that may shift state evolution toward the reference regime. In this manuscript, "signature information" denotes a structured, intervention-specific input intended to couple to a diagnostically identified disturbance pattern. "Context information" denotes endogenous or relational influences (expectation, meaning, care context) that also change state.

Operationally, deviation is approximated by preregistered indication-specific proxies ("D-proxies"): composite symptom scales, objective burden markers, and regulation/coherence metrics derived from time series or biomarker clusters. The evaluation criterion is strictly empirical: under matched context, a signature hypothesis is supported only if preregistered outcome contrasts show the predicted pattern and pass falsification rules.

4. Potentization as A Signature Hypothesis

Potentization is interpreted as a procedure intended to reduce material degrees of freedom while preserving or amplifying a structured intervention signature. This is a conditional claim: absence of molecules does not imply efficacy. The framework remains mechanism-open and is evaluated entirely through prediction-driven tests.

5. Materials and Methods

5.1. Why Diagnostic Precision Should Moderate Specific Effects

If an intervention is diagnosis-specific, imperfect diagnostic classification dilutes the target-matching proportion within the treated group. As diagnostic certainty decreases, misclassification increases, attenuating the observed contrast between a diagnosis-congruent protocol and placebo even if a true diagnosis-specific effect exists. Therefore, under a genuine diagnosis-specific signature add-on, the congruent–placebo contrast should increase as the probability of correct classification increases.

5.2. Core Prediction with Outcome-Based Effect Definition

Let the preregistered primary outcome be $Y = \Delta D^*$ (change in a preregistered composite deviation proxy). Define the diagnosis-specific add-on effect at diagnostic certainty level S_{diag} as the conditional outcome contrast:

$$E_{\text{spec}}(S_{\text{diag}}) = \mathbb{E}[Y \mid Qc, S_{\text{diag}}] - \mathbb{E}[Y \mid P, S_{\text{diag}}].$$

The central hypothesis is diagnostic moderation:

$$E_{\text{spec}}(S_{\text{diag}}) = f(S_{\text{diag}}), \qquad f'(S_{\text{diag}}) > 0,$$

under a matched-context design with a diagnosis-incongruent mismatch arm.

5.3. Operationalizing Diagnostic Certainty $\boldsymbol{S_{diag}}$ (Measurement Model + Proxy Fallback)

To avoid circularity, S_{diag} is preregistered and computed independently of outcomes.

Primary (recommended): Measurement-model score as estimated correctness probability. Let $Z_i \in \{1, \dots, K\}$ be the latent true diagnostic class for individual i. Let R_{ij} be the label assigned by rater j under a prespecified rubric, blinded to outcomes. Optionally, let T_{im} be auxiliary test features (e.g., prespecified biomarker-pattern indicators) measured at baseline. A preregistered Bayesian latent-class model estimates posterior class probabilities:

$$S_{\text{diag},i} \equiv \hat{p}_i = \max_k P(Z_i = k \mid R_{i1:J}, T_{i1:M}),$$

with priors and identifiability constraints prespecified. The monotone hypothesis is then explicitly about $\hat{p}_i$ strata (e.g., quartiles) or a continuous $\hat{p}_i$ moderator.

Fallback (if measurement model is infeasible): Proxy strata. If only agreement metrics are available, S_{diag} is defined as a preregistered proxy stratum (e.g., Fleiss' κ category computed from $J \geq 3$ raters). In that case, the hypothesis is stated as: *the expected add-on contrast increases across preregistered proxy strata*, with explicit sensitivity analyses for confounding (see below).

Confounding control and sensitivity. Primary analyses adjust for preregistered covariates (e.g., baseline burden D_0^*, severity/stage, co-treatment indicators). Randomization is stratified by baseline severity. A preregistered sensitivity analysis tests whether S_{diag} is correlated with baseline burden and whether moderation persists after adjustment.

5.4. Equivalence-Defined Decision Rules

Define Y on a common scale (e.g., standardized change score) so that equivalence margins are interpretable. Let SMD_{AB} denote the standardized mean difference on Y between arms A and B.

Equivalence margin for mismatch arm. Preregister an equivalence margin $\delta > 0$ for the mismatch vs placebo contrast:

$$H_{0,\text{neq}}\colon\ \left|\text{SMD}_{Qm,P}\right| \geq \delta \quad \text{vs.} \quad H_{1,\text{eq}}\colon\ \left|\text{SMD}_{Qm,P}\right| < \delta.$$

Test via TOST (two one-sided tests) or via Bayesian ROPE with the same δ.

Minimal relevance threshold for congruent arm. Preregister a minimal relevant effect $\delta_{\min}$ for the congruent vs placebo contrast:

$$\text{SMD}_{Qc,P} \geq \delta_{\min},$$

in addition to statistical evidence (frequentist or Bayesian). If a validated MCID exists for the chosen clinical scale, $\delta_{\min}$ can be defined in that native unit.

Primary decision rule. Support for a diagnosis-specific add-on requires all of: (i) $\text{SMD}_{Qc,P} \geq \delta_{\min}$ with preregistered statistical evidence, (ii) equivalence of Qm and P (i.e., $\left|\text{SMD}_{Qm,P}\right| < \delta$ under TOST/ROPE), and (iii) a positive monotone trend of $(Qc - P)$ with increasing S_{diag} (preregistered trend test or Bayesian model comparison).

Falsification rule. Reject the signature add-on if any of: (i) $\text{SMD}_{Qc,P} < \delta_{\min}$ (or fails preregistered evidence threshold), (ii) Qm is non-equivalent to P (context/expectancy artifact likely), (iii) the monotone trend criterion is not met, or reverses direction.

5.5. Three-Arm Matched-Context Design

A preregistered three-arm study is proposed:

1. Diagnosis-congruent protocol (Qc),
2. Placebo under identical context (P),
3. Diagnosis-incongruent mismatch protocol (Qm) under identical context.

Matched context is implemented by a double-dummy procedure with identical dosing schedules, identical visit frequency and interaction time, identical packaging/ritual cues, and standardized scripts for participant contact across arms. In the mismatch arm, the intervention is diagnosis-incongruent but procedurally identical to the congruent arm (same carriers, packaging, dosing schedule, interaction time, and scripted communication), so that only diagnostic congruence varies. Expectancy and credibility are measured with a prespecified instrument and included as preregistered covariates or balance checks. Blinding integrity is assessed using preregistered blinding-check items.

5.6. Worked Example: Acute Otitis Media (AOM)

To make D^* fully concrete, we provide a worked example for *acute otitis media in children*, a condition represented among individualized homeopathy RCTs (e.g., Jacobs et al. 2001) (Jacobs et al., 2001). This section is a *specification template* for preregistration, not a claim about clinical efficacy.

Population: children with clinician-diagnosed AOM meeting prespecified inclusion criteria.

Baseline assessment window: day 0 (two baseline measurements if biomarker-pattern stability is used).

Follow-up window: days 1–7 (daily symptom capture), with a prespecified primary timepoint (e.g., day 3).

Clinical component D_{clin}: preregister a validated symptom severity scale (instrument name, scoring, and direction reproduced exactly as published). Define D_{clin} as the standardized score at the primary timepoint.

Objective component D_{obj}: prespecify an objective marker, e.g., analgesic rescue use (dose count) and/or blinded otoscopic severity coding.

Regulation/coherence component D_{reg} (optional): if a preregistered time-series proxy is available (e.g., sleep fragmentation), define it with a fixed algorithm.

Composite:

$$D^* = w_1 D_{\text{clin}} + w_2 D_{\text{obj}} + w_3 D_{\text{reg}},$$

with (w_1, w_2, w_3) fixed a priori and sensitivity-tested (e.g., alternative weight sets preregistered as robustness checks).

Primary endpoint: $Y = \Delta D^* = D^*_{\text{baseline}} - D^*_{\text{primary timepoint}}$ (direction preregistered).

Margins: define δ (equivalence) and $\delta_{\min}$ (minimal relevance) in SMD units (e.g., $\delta = 0.20$, $\delta_{\min} = 0.20$) or in native units if a referenced MCID exists.

Missing data: preregister mixed-effects model with MAR assumption + multiple imputation sensitivity; define drop-out handling and protocol deviations.

Preregistration checklist (worked example). (i) Eligibility and diagnostic rubric; (ii) how S_{diag} is computed (measurement model or proxy); (iii) full endpoint definition and scoring; (iv) δ and δ_{min}; (v) analysis plan (models, covariates, multiplicity); (vi) blinding/expectancy measurement; (vii) falsification criteria.

5.7. Validation Studies: In Silico Suite + External Consistency Check

This manuscript reports no new patient-level clinical data. It reports (i) an in silico validation suite and (ii) a secondary evidence check based on published meta-analytic summary estimates.

Study 1A (Null): Type-I error control. Simulate data under $\beta_{\text{spec}} = 0$ (no diagnosis-specific add-on), including a scenario in which baseline burden correlates with S_{diag}. Verify that the preregistered moderation test controls false positives.

Study 1B (Alternative): Power sweeps. Simulate diagnosis-moderated add-on via $M_i \sim \text{Bernoulli}(S_{\text{diag},i})$ in Qc and $M_i = 0$ in P and Qm; evaluate power for $T \times S_{\text{diag}}$ across N and effect parameters.

Study 1C (Stress test): Blinding/expectancy leakage assay. Add an expectancy pathway that increases outcomes in both active-procedure arms (Qc and Qm) under partial unblinding. Report the exceedance rate $P(|\text{SMD}_{Qm,P}| > \delta)$ as a heuristic flag for expectancy artifacts (no formal TOST/ROPE equivalence test is implemented in the simulation suite); in an empirical preregistration, equivalence would be evaluated via TOST or a Bayesian ROPE decision rule using the same δ.

Study 2 (External consistency check). We do not treat Mathie et al. (2014) as diagnostic certainty. Instead, we use their risk-of-bias–based designation of "reliable evidence" as a study-level reliability moderator, S_{qual}. A trial is "reliable evidence" if it is A-rated (low risk of bias in all seven domains), or B1-rated with uncertainty restricted to domains IV–VI only (Mathie et al., 2014). We contrast the pooled odds ratio across all 22 meta-analysable trials with the sensitivity estimate restricted to the three "reliable evidence" trials, as reported by the authors (Mathie et al., 2014). We explicitly note that this is a quality-stratification pattern, not a test of diagnostic moderation.

6. Results

6.1. Study 1 (In Silico): Type-I, Power, And Leakage Assay

We implemented the simulation suite (Appendix code).

Null (type-I error). Under the null ($\beta_{\text{spec}} = 0$) with baseline–S_{diag} correlation and baseline adjustment, the Monte Carlo type-I error for the $T \times S_{\text{diag}}$ moderation test was approximately 0.051 (1,000 replicates; $\alpha = 0.05$).

Alternative (power sweeps). Under an alternative with a diagnosis-moderated add-on (parameters in Appendix), power increased with sample size (illustrative values: $N = 100 \rightarrow 0.11$, $N = 200 \rightarrow 0.238$, $N = 400 \rightarrow 0.44$; 500 replicates each).

Leakage stress test (assay behavior). Under expectancy leakage, the mismatch arm shifts upward relative to placebo. Using the preregistered *equivalence margin concept* as an operational flag, the mismatch contrast exceeded the illustrative margin $\delta = 0.20$ in magnitude in approximately 0.683 of runs (1,000 replicates; mean $\text{SMD}_{Qm,P} \approx 0.276$), whereas under the null it exceeded δ in about 0.275 of runs (mean $\text{SMD}_{Qm,P} \approx 0.010$). This is reported as an exceedance-rate heuristic (not a formal equivalence test): leakage increases the frequency with which $|\text{SMD}_{Qm,P}|$ exceeds the preregistered margin δ; in a real study, equivalence would be assessed via TOST (or Bayesian ROPE) using the same δ.

6.2. Study 2: External Consistency Check Using Risk-of-Bias–Based Reliability

Across 22 meta-analysable RCTs of individualised homeopathy, Mathie et al. (2014) report OR = 1.53 (95% CI 1.22–1.91) (Mathie et al., 2014). Restricting to the three trials designated as "reliable evidence" yields OR = 1.98 (95% CI 1.16–3.38) (Mathie et al., 2014). The three trials meeting these strict criteria are Jacobs et al. (1994), Jacobs et al. (2001), and Bell et al. (2004) (Jacobs et al., 1994; Jacobs et al., 2001; Bell et al., 2004). We further note that Mathie et al. report no statistically significant difference between the "reliable" and "non-reliable" subsets in their sensitivity comparison (reported p-value: 0.42), underscoring that this check is a consistency sanity test rather than confirmatory evidence for diagnostic moderation (Mathie et al., 2014).

Secondary evidence check using published meta-analytic estimates (Mathie et al., 2014)

Stratum	OR	95% CI
All meta-analysable trials (n=22)	1.53	1.22–1.91
"Reliable evidence" trials (n=3)	1.98	1.16–3.38

7. Discussion

7.1. Placebo as Endogenous Information Regulation

Placebo and context effects are endogenous state updates and therefore support the broad informational framing while simultaneously demanding strict separation between context information and any diagnosis-specific add-on. Because diagnoses can shape expectations, a mismatch arm under matched context and an equivalence-defined criterion are necessary for credible discrimination.

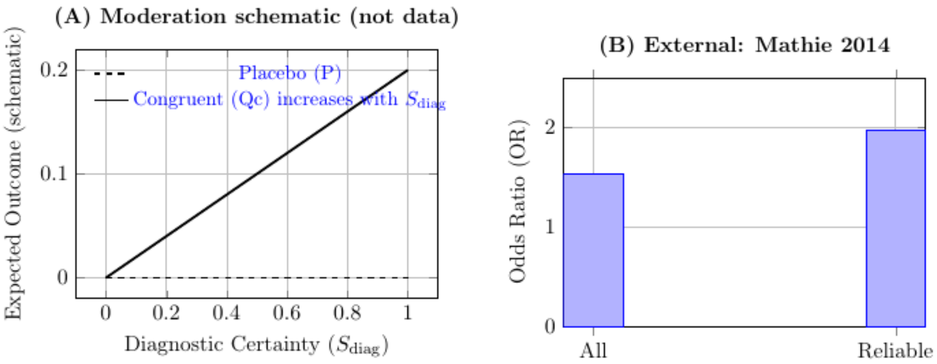

Figure 1. Design logic and external sensitivity pattern.

(A) Schematic: under a diagnosis-specific add-on, the congruent–placebo separation increases with preregistered S_{diag}.

(B) External consistency check: restricting to risk-of-bias "reliable evidence" trials changes the pooled OR from 1.53 to 1.98; this is a study-quality sensitivity pattern and does not test diagnostic-certainty moderation.

7.2. *Counterpositions and Skeptical Constraints*

Skeptical assessments emphasize weak robustness under strict quality filters and the implausibility of ultra-high dilutions under substance-based causal models (EASAC 2017; Grams 2019). If the skeptical null is universally correct, then $f(S_{\text{diag}}) = 0$ across strata and the three-arm design yields no add-on. A stable positive slope under preregistration is a specific empirical deviation from that null.

7.3. *Why the Measurement Model Matters*

Without a measurement model, "diagnostic certainty" is only a proxy. By preregistering S_{diag} as an estimated correctness probability $\hat{p}_i$ (or by explicitly restricting inference to proxy strata), the moderation claim becomes operationally testable rather than interpretive.

7.4. *Ethical Framing*

This manuscript proposes a research framework, not a clinical guideline. Hypotheses with unclear evidential strength should not displace better-supported, lower-risk options. Ethical evaluation is anchored in diagnostic accuracy, demonstrable efficacy, transparent uncertainty management, and avoidance of preventable harm.

8. Limitations

The framework is mechanism-open; it does not specify a physical carrier for ultra-high potencies. The external consistency check is a quality-stratification pattern and is not evidence for diagnostic moderation. The approach stands or falls solely with preregistered moderator tests and mismatch-controlled discrimination, including equivalence-defined falsification.

9. Conclusion

This manuscript does not argue efficacy; it specifies a preregisterable discrimination pattern that makes the controversy experimentally decidable. If a diagnosis-specific signature add-on exists beyond context effects, the congruent–placebo contrast must increase monotonically with preregistered diagnostic certainty under matched context, while the diagnosis-incongruent mismatch arm remains equivalent to placebo within a preregistered margin. The framework is empirically strict: it stands or falls by endpoints, blinding/expectancy controls, equivalence margins, and falsification rules.

Abbreviations:

The following abbreviations are used in this manuscript.

Qc	Diagnosis-congruent protocol arm
Qm	Diagnosis-incongruent mismatch protocol arm
P	Placebo arm (matched context)
D	Deviation / incoherence measure
D^*	Preregistered composite deviation proxy
E_{spec}	Diagnosis-specific add-on effect
S_{diag}	Diagnostic certainty score
SMD	Standardized mean difference
TOST	Two one-sided tests (equivalence testing)

Data Availability Statement:

Simulation code is provided in the Appendix. The external consistency check uses published meta-analytic summary estimates from Mathie et al. (2014), which are publicly available in the referenced literature.

References

Linde, K., Clausius, N., Ramirez, G., Melchart, D., Eitel, F., Hedges, L. V., & others. (1997). Are the clinical effects of homoeopathy placebo effects? A meta-analysis of placebo-controlled trials. *The Lancet, 350*(9081), 834–843. https://doi.org/10.1016/S0140-6736(97)02293-9

Shang, A., Huwiler-Müntener, K., Nartey, L., Jüni, P., Dörig, S., Sterne, J. A. C., & others. (2005). Are the clinical effects of homoeopathy placebo effects? Comparative study of placebo-controlled trials of homoeopathy and allopathy. *The Lancet, 366*(9487), 726–732. https://doi.org/10.1016/S0140-6736(05)67177-2

Mathie, R. T., Lloyd, S. M., Legg, L. A., Clausen, J., Moss, S., Davidson, J. R. T., & Ford, I. (2014). Randomised placebo-controlled trials of individualised homeopathic treatment: Systematic review and meta-analysis. *Systematic Reviews, 3*, 142. https://doi.org/10.1186/2046-4053-3-142

Mathie, R. T., Ramparsad, N., Legg, L. A., Clausen, J., Moss, S., Davidson, J. R. T., Messow, C.-M., & McConnachie, A. (2017). Randomised, double-blind, placebo-controlled trials of non-individualised homeopathic treatment: Systematic review and meta-analysis. *Systematic Reviews, 6*, 63. https://doi.org/10.1186/s13643-017-0445-3

European Academies' Science Advisory Council (EASAC). (2017). *Homeopathic products and practices: Assessing the evidence and ensuring consistency in regulating medical claims in the EU*. https://easac.eu/publications/details/homeopathic-products-and-practices

Grams, N. (2019). Homeopathy—Where is the science? A current inventory on a pre-scientific artifact. *EMBO Reports, 20*(7), e47761. https://doi.org/10.15252/embr.201947761

Hamre, H. J., Glockmann, A., von Ammon, K., Riley, D. S., Kienle, G. S., & Kiene, H. (2023). Efficacy of homoeopathic treatment: Systematic review of meta-analyses of randomised placebo-controlled homoeopathy trials for any indication. *Systematic Reviews, 12*, 191. https://doi.org/10.1186/s13643-023-02313-2

Banerji, P., Campbell, D. R., & Banerji, P. (2008). Banerji protocols for the treatment of cancer with homeopathic medicine. *Oncology Reports, 20*(1), 69–74. https://doi.org/10.3892/or.20.1.69

Olaku, O., Zia, F., Santana, J. M., & White, J. D. (2013). The National Cancer Institute Best Case Series Program: A summary of cases of cancer patients treated with the Bangladesh and India homeopathy project. *Integrative Cancer Therapies, 12*(2), 156–164. https://doi.org/10.1177/1534735412473641

Jacobs, J., Jimenez, L. M., Gloyd, S. S., Gale, J. L., & Crothers, D. (1994). Treatment of acute childhood diarrhea with homeopathic medicine: A randomized clinical trial in Nicaragua. *Pediatrics, 93*(5), 719–725.

Jacobs, J., Springer, D. A., & Crothers, D. (2001). Homeopathic treatment of acute otitis media in children: A preliminary randomized placebo-controlled trial. *The Pediatric Infectious Disease Journal, 20*(2), 177–183.

Bell, I. R., Lewis, D. A., Jr., Brooks, A. J., Schwartz, G. E., Lewis, S. E., Walsh, B. T., & Peterson, C. M. (2004). Improved clinical status in fibromyalgia patients treated with individualized homeopathic remedies versus placebo. *Rheumatology, 43*(5), 577–582.

Supplementary Material: Simulation Code (Suite)

Listing 1: Python simulation code for the proposed validation suite.

```python
import numpy as np
import pandas as pd
import statsmodels.formula.api as smf
from scipy import stats

def one_sim(rng, N=200, beta_spec=0.0, beta_context=2.0, sigma=2.5,
            rho_bs=0.4, beta_base=1.0,
            blinding_leak=False, leak_prob=0.5, beta_expect=2.0):

    S = rng.uniform(0,1,N)
    z1 = rng.normal(size=N)
    D0 = (z1 - z1.mean())/z1.std()
    S = 0.5*S + 0.5*stats.norm.cdf(rho_bs*D0 + np.sqrt(1-rho_bs**2)*rng.normal(size=N))
    arm = rng.integers(0,3,size=N)
    M = np.zeros(N)
    is_Qc = arm==1
    M[is_Qc] = rng.binomial(1, S[is_Qc])

    E = np.zeros(N)
    if blinding_leak:
        active = (arm!=0)
        leak = rng.binomial(1, leak_prob, size=N)*active
        E = leak*(0.5+0.5*S)

    eps = rng.normal(0,sigma,N)
    Y = 10.0 + beta_context + beta_base*D0 + beta_spec*M + beta_expect*E + eps

    df = pd.DataFrame({"Y":Y,"S":S,"arm":arm,"D0":D0})
    df_qc_p = df[df["arm"].isin([0,1])].copy()
    df_qc_p["T"] = (df_qc_p["arm"]==1).astype(int)
    model = smf.ols("Y ~ T + S + D0 + T:S", data=df_qc_p).fit()
    p_int = model.pvalues.get("T:S", np.nan)

    y_qm = df[df["arm"]==2]["Y"].to_numpy()
    y_p  = df[df["arm"]==0]["Y"].to_numpy()

    if len(y_qm)>1 and len(y_p)>1:
```

```
        smd  = (y_qm.mean()-y_p.mean())/np.sqrt((y_qm.var(ddof=1)+y_p.var(ddof=1))/2)
    else:
        smd  = np.nan

    return p_int, smd

def sim_suite(nsim=1000, seed=12345, delta=0.20, **kwargs):
    rng = np.random.default_rng(seed)
    pints=[]; smds=[]; exceed=[]
    for _ in range(nsim):
        p_int, smd = one_sim(rng, **kwargs)
        pints.append(p_int); smds.append(smd)
        exceed.append(abs(smd) > delta)

    pints=np.array(pints); smds=np.array(smds, dtype=float); exceed=np.array(exceed, dtype=bool)

    return {
        "moderation_reject_rate": float(np.nanmean(pints<0.05)),
        "qm_vs_p_mean_smd": float(np.nanmean(smds)),
        "qm_vs_p_exceed_delta_rate": float(np.nanmean(exceed))
    }

print("Study 1A (Null):",
      sim_suite(N=200, beta_spec=0.0, blinding_leak=False, nsim=1000, delta=0.20))

for N in [100,200,400]:
    out = sim_suite(N=N, beta_spec=3.0, blinding_leak=False, nsim=500, delta=0.20)
    print("Study 1B (Alt) N=",N, out)

print("Study 1C (Leakage):",
      sim_suite(N=200, beta_spec=3.0, blinding_leak=True,
                leak_prob=0.5, beta_expect=2.0, nsim=1000, delta=0.20))
```