

Independent Statistical Review Brief

SEC-CDP-001 Pilot v0.4.1 | Brief v1.0 | Package v1.3 | 11 October 2026

Prepared for an independent statistician on behalf of Thomas Vargo / Aegis Solis. Please assess whether the proposed feasibility pilot has a coherent target, scoring policy and analysis plan, and identify what must change before registration or recruitment. This is a draft review request, not evidence of reviewer appointment, ethics approval or an executed study.

Design and intended use

The proposal randomizes up to 60 adults approximately equally between conspicuous-error exposure and corrected-output Control, using concealed blocks of four or six within two baseline-expertise strata. After eight exposure outputs, every participant reviews the same fixed bank of 48 challenge items: 24 targets and 24 benign items across three families. Each item has a 60-second maximum, with voluntary terminal submission. Six cyclic order lists support descriptive early, middle and late contrasts. The bank, interface, order lists and executable analysis are not yet delivered or validated.

The pilot is intended to assess feasibility, measurement and analysis behavior. It is not a confirmatory efficacy test and cannot establish AGI alignment, human indispensability or a safety benefit. Detection and benign false alarms must be reported together. The full-bank target is the randomized cohort and observed bank, with equal family weights and a common empirical baseline-expertise distribution.

Availability decisions to assess

Output	Minimum cell-count condition	Declared limitation
Main full-bank point summary	At least 1 randomized participant per arm in every required expertise stratum	Missing outcomes can still prevent point identification.
Proposed full-bank bootstrap interval	At least 5 per arm in every required expertise stratum	Coverage still requires independent review and simulation.
Descriptive position-band contrasts	At least 6 per arm in every included expertise stratum under the six-list cycle	No band interval is selected; missing outcomes remain relevant.

There is no stratum recruitment floor. Unavailability is accepted in advance; no replacement arm contrast or interval is selected. Empty required cells also make the selected composite bound formula unavailable. Please assess whether this is useful for a 60-person pilot and whether any redesign should be frozen before recruitment. Do not choose an alternative after observing outcomes.

Questions for the reviewer

- 1. Target and allocation:** confirm that the finite-bank, recruited-cohort estimand matches the estimator, blocked allocation and common expertise weights. Assess the implication of convenience recruitment and repeated items. Distinguish this exploratory pilot from the later confirmatory randomization-based procedure in Crosswalk sections 14.2 and 14.4.
- 2. Uncertainty:** assess the proposed 10,000 evaluator bootstrap resamples within arm and expertise stratum, with each evaluator's full item vector retained. Specify simulations for the actual allocation, list assignment, cell occupancy, response heterogeneity and missingness. Evaluate coverage, bias, undefined estimates and Monte Carlo error; a five-person threshold is not a validation result.
- 3. Missing outcomes and scoring:** assess assigned-denominator identification bounds, known-event retention and the masking of every interrupted item in the optional complete-item model. Check whether any intended interpretation requires additional missingness assumptions. Accepted retracted flags remain cumulative events; late flags are archived and excluded. The reference function trusts supplied event ordering and adjudication.
- 4. Time and order:** assess voluntary submission as terminal, no-detection-through-60 scoring for complete misses, detection incidence and restricted mean time without detection. Separate actual time spent from the task-policy endpoint. Review early technical loss assumptions and equal-list weighting for bands, including empty-cell behavior and repeated-exposure carryover.
- 5. Feasibility and progression:** assess whether 60 randomized people can answer the operational questions, the proposed completion/loss/agreement criteria, and the implications of no post-randomization replacement. Effect sign and statistical significance are not progression criteria. Review uncertainty and missingness treatment for the feasibility thresholds and the separation from later confirmatory sample-size planning.

Requested written assessment

Identify blocking, major and minor issues with locations and proposed corrections. State which targets and reports are defensible, which remain assumption-dependent or unavailable, and what analysis implementation and simulations are required. Recommend any design changes before registration, with their consequences for the target and resources. Record reviewer identity, relevant expertise, independence, conflicts, reviewed versions and review date; this brief supplies no completed sign-off.

Reading order and remaining gates

Start with package START_HERE and Pilot sections 4 and 7-10. Consult Crosswalk sections 13-14 for the broader research framework; review the scorer and verifier only for their stated code scope. The full supplied reviews and exact prior package are in 06_BASELINES. All eight pilot gates remain open, including ethics, materials, interface/order lists, allocation, analysis validation, statistical review and registration/execution approval. Infrastructure review is a separate engineering task; no live system is needed for this fixed-bank proposal.