

# New Investment & Measurement Assurance

An incumbent is evidence to inspect, not an automatic rejection

A white-space investment and measurement-integrity case. It separates programme discovery, functional equivalence, local applicability and marginal value while preserving the costs of both duplicate funding and mistaken rejection.

## WHITE-SPACE REVIEW & MEASUREMENT INTEGRITY

### BLUF / BOTTOM LINE UP FRONT

Verify what already works. Price what the proposed change would add.

### ANALYTICAL FINDING

A catalogue hit cannot distinguish a duplicate from a useful integration. A rejection fraction is not a verified error rate. Smaller, non-additive credit still needs a counterfactual before it becomes incremental value.

### CONTRIBUTION IN THE SOURCE RECORD

Provisional registry-level hoop for authored analysis, revision and disclosure. Programme execution, approval and realized impact remain separate.

### EVIDENCE STATUS

Exact review and business-case originals remain unrecovered. A related planning workbook does not authenticate their submission or outcomes.

BLUF: analytical inference. Source contribution: historical record. Client and internal identities anonymized.

STRUCTURED ABSTRACT

# Evaluate the proposed increment before accepting a novelty label

[AI] Evidence of an incumbent should change the comparison, not end it. The right decision may be reuse, local integration, a new build or a bounded hold.

| Element             | Decision-relevant conclusion                                                                                                                                                                                                                                          |
|---------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Context [HR]        | A planning lead reportedly produced white-space ideas, programme coverage maps, skeptical reviews and a smaller measurement/execution proposal. The selected records describe document work, not authenticated submission, funding or deployment.                     |
| Objective [AI]      | Distinguish finding an incumbent from proving equivalent coverage, and identify what the proposed change adds relative to the available continuation. Price mistaken rejection as well as duplicate investment.                                                       |
| Approach [RC/AI]    | Reconcile screen populations, status labels and estimate bases privately. Examine evidence independence, physical-versus-recorded outcomes, incremental attribution and the availability of stronger low-cost alternatives.                                           |
| Findings [AI]       | A reported rejection fraction is not a verified error rate. Live, in-development and locally usable are different states. A non-additive share of another programme’s planned benefit does not identify an enabler’s causal contribution.                             |
| Decision value [PD] | The sponsor defines admissible actions and error costs; programme owners confirm scope and access; Finance approves the valuation basis; an independent reviewer validates the evidence contract. Keep unverified ideas visible without inventing novelty or benefit. |
| Implications [PD]   | Use the Novelty-to-Increment Evidence Contract (NIEC). In the constructed benchmark, better triage can lower errors yet fail its price test. More complete evidence can outperform the tested review, but still requires calibrated cost, accuracy and authority.     |

[PD] The first release is a versioned comparison: current programme, remaining eligible gap, proposed action, incremental cost and the evidence that could reverse the decision. This case proposes no automatic rejection, funded pilot or new customer policy.

[HR] historical record; [RC] recalculated; [AI] analytical inference; [PD] proposed design. Two derivative records support bounded authorship claims. All historical numerical coordinates remain private.

EXHIBIT 1 / DECISION-ERROR DENOMINATOR

# A rejection count does not measure whether the review was right

[AI] A screening process can reject many proposals and still destroy valuable options. The verdict is an output to validate, not its own ground truth.

| Independent construction           | World A                        | World B                                   |
|------------------------------------|--------------------------------|-------------------------------------------|
| Original candidates                | 20                             | 20                                        |
| Candidates rejected                | 12                             | 12                                        |
| Useful candidates wrongly rejected | 0                              | 8                                         |
| Meaning of the same observed count | Potentially accurate screening | Material false rejection remains possible |

[PD] These invented worlds share the same reported rejection count. In World A the rejected ideas are duplicates or unviable. In World B some rejected ideas contain useful, currently permissible changes. The count alone cannot distinguish them. To estimate a false-positive rate, specify which class is positive and independently establish the relevant truth denominator; selected proposed ideas do not supply a population of all opportunities.

[HR/AI] The source describes skeptical, citation-backed reviews that led to rejection, rescoping and retention. That is evidence of an analytical process, not an independently audited confusion matrix. Multiple agents can share the same catalogue, source omissions, retrieval pathway or interpretation. Requiring a URL improves traceability but does not establish independence or equivalence.

[AI] A deliberate bias toward rejection may be appropriate under some agreed costs. It is not intrinsically conservative: a false rejection can forgo an integration benefit, delay a needed measurement repair or preserve an expensive workaround. The decision owner must specify which mistakes matter and how to detect them.

AUDIT BOTH SIDES OF THE VERDICT

[PD] Independently sample retained and rejected proposals, keep unavailable references unresolved, and review the original scope without revealing the prior label. Report false rejection, unsupported funding and unresolved coverage on their own original denominators.

[PD] Independent counterexample, not an estimated error rate for the historical review. A high agreement rate between dependent reviewers is not independent validation.

EXHIBIT 2 / FUNCTIONAL EQUIVALENCE

# Existing technology can leave a useful integration gap

[AI] A programme name answers an existence question. An investment decision requires a comparison of functions, eligible exposure, action rights and delivery conditions.

| Equivalence dimension        | Question the review must resolve                                                          |
|------------------------------|-------------------------------------------------------------------------------------------|
| Function and outcome         | Does the incumbent perform the same action on the same failure mechanism?                 |
| Eligibility and channel      | Can it serve the proposed population, including exceptions and partner-operated paths?    |
| Version and operating status | Is it deployed, in development, optional or unavailable at this decision date?            |
| Authority and capacity       | Can the owner lawfully execute the action with accepted resources and integration rights? |
| Increment and cost           | What useful change remains after reuse, and what does it cost beyond current operations?  |

[AI] A shared title or defect bucket can conceal different mechanisms. A capability may exist centrally while a local interface, prerequisite, adoption path or reference measurement remains missing. Conversely, a locally absent record may simply reflect incomplete search. Neither a catalogue hit nor a failed search resolves these alternatives.

[HR/AI] The source usefully challenged its own initial catalogue and revised candidate descriptions against incumbent evidence. However, reported live and in-development findings cannot be merged into a claim that all rejected ideas were already operating. The surviving measurement work also faces an integration question: an existing observability system may absorb it, complement it or leave a material gap.

CHANGE THE COMPARATOR BEFORE THE LABEL

[PD] Compare new build, reuse, integration and bounded deferral against the same existing continuation. Preserve a scoped difference statement and owner response. Do not assign individual fault, duplicate-benefit deductions or zero marginal value from a shared programme name alone.

[AI] Absence from searched systems is not proof of absence everywhere. Programme existence is neither a causal estimate nor automatic permission to deploy or integrate.

EXHIBIT 3 / MEASUREMENT AND EXECUTION

# Prove the action separately from the effect it is meant to cause

[AI] A complete measurement record can improve accountability while leaving physical service unchanged. Each link in the claim requires its own reference.

| Evidence layer                  | What it establishes                      | What remains open                              |
|---------------------------------|------------------------------------------|------------------------------------------------|
| Action named and owner recorded | A stated commitment                      | Acceptance, authority and capacity             |
| Task marked complete            | A recorded workflow state                | Actual execution and eligible reach            |
| Independent execution evidence  | The specified action occurred            | Its effect versus continuation                 |
| Leading indicator moves         | A measured association at a chosen clock | Capture changes, confounding and lag selection |
| Mature outcome differs          | An observed outcome contrast             | Causal attribution and avoidable value         |

[HR/AI] The proposed mechanism would make commitments, execution and outcomes more auditable. That is a credible analytical objective. Yet missing end-to-end evidence does not establish that no action occurred; a later indicator change does not validate execution; and a maturity adjustment does not supply the missing counterfactual. The source’s reported operational deterioration after a scorecard change is also compatible with other simultaneous changes.

[AI] Reclassification can improve a metric’s meaning while moving counts among owners without preventing a customer failure. Better capture can increase the recorded defect count even when underlying service is unchanged. A new ledger therefore needs an independent physical reference and stable original eligibility, not only reconciled categories or cleaner dashboards.

INSTRUMENT THE CLAIM, THEN TEST THE RESPONSE

[PD] Freeze the action, population, deadline and measurement version. Verify execution independently; then test authorized response on original cohorts with mature outcomes and full work costs. Keep reporting repair, execution coverage and physical benefit in separate ledgers.

[HR/AI] No historical implementation or causal operating improvement is established. The proposed shadow model evaluates advice, not customer service or a production measurement deployment.

EXHIBIT 4 / INCREMENTAL ECONOMIC VALUE

# A smaller share of a parent plan is still not an identified increment

[AI] Calling a benefit non-additive avoids one bookkeeping error. It does not determine how much value an enabler actually causes.

| Independent outcome construction | Without proposed enabler | With enabler | Incremental improvement |
|----------------------------------|--------------------------|--------------|-------------------------|
| World 1                          | 80                       | 80           | 0                       |
| World 2                          | 80                       | 70           | 10                      |
| World 3                          | 80                       | 90           | -10                     |

[PD] Lower outcomes are better in this invented example. All worlds can coexist with the same parent plan and the same accounting share assigned to the enabler. The share cannot distinguish no effect, improvement or harm. Only an appropriate comparison identifies the increment. An enabler may be essential without a standalone benefit; that requires evidence about the joint package, not a percentage selected from its headline.

[HR/RC] The source revised an optimistic leverage claim and described a smaller share of the parent trajectory. This is useful disclosure, but neither the revised ratio nor the narrower range validates the effect, timing, coverage or conversion to avoidable cost. The proposed staffing envelope is an estimate, not an observed operating expense or accepted capacity commitment.

[AI] Annualizing a selected week does not establish a mature annual benefit. Concessions and loss categories may refer to overlapping events, and one customer can appear more than once. Do not add them into unique customers helped without an identity and timing bridge. A physical improvement may also fail to release cash if the affected expense remains fixed.

PRICE THE CHANGE RELATIVE TO CONTINUATION

[PD] Retain build, integration, review, maintenance and displaced work. Separate avoided cost, recognized benefit and cash release. Report the conditions that would reverse the investment, including zero increment or an incumbent capable of delivering it more cheaply.

[PD] Independent counterfactual proof; no historical quantity calibrates it. [RC] Conflicting rate/count conversions, annualizations and ratio corners remain in the private arithmetic ledger.

EXHIBIT 5 / REVIEW AND VERSION LINEAGE

# Several review rounds do not become one validated funnel

[AI] A decision trail is useful only when each transition refers to the same candidate, scope, evidence date and disposition rule.

| Record to preserve                     | Failure prevented                                                       |
|----------------------------------------|-------------------------------------------------------------------------|
| Stable proposal and version ID         | Treating a rewritten integration as the original rejected build.        |
| Search frame and accessible sources    | Promoting an unobserved capability into verified absence.               |
| Incumbent scope and operating date     | Equating planned development with live local availability.              |
| Decision reason and reversal condition | Turning a provisional verdict into an irreversible global label.        |
| Submission and approval artifact       | Treating a document in submission format as a sent or accepted request. |

[HR/AI] The selected records describe idea generation, catalogue comparison, skeptical review, a later reduction and final document shaping. These rounds do not necessarily share one denominator or a single proposal grain. Linking them into a smooth funnel would imply a lineage that the current evidence does not authenticate. Preserve each population and map revisions explicitly.

[AI] A source can contain both a retrospective claim of submission and an explicit absence of filing or approval evidence. Neither should silently replace the other. Use a typed status: prepared, reportedly sent, authenticated sent, reviewed, approved and deployed. A version timestamp or generated-file metadata does not establish the decision event.

MAKE REVERSAL POSSIBLE

[PD] Store the cited scope, contrary evidence, unresolved questions and the owner authorized to reopen a verdict. A later production release, changed eligibility or cheaper integration can rationally change a decision. That is a new evidence state, not proof that the earlier reviewer was dishonest.

[HR/AI] Exact review outputs, candidate mappings and decision artifacts require authentication. Internal analytical reproducibility is not an external independent audit of historical verdicts.

EXHIBIT 6 / CONTRIBUTION AND AUTHORITY

# Credit the analytical revision without claiming the programme result

[AI] The defensible contribution is the evidence and decision work described in the records. It does not inherit the performance of the programmes cited or proposed.

| Activity or claim                        | Bounded attribution                                                                       |
|------------------------------------------|-------------------------------------------------------------------------------------------|
| Framing, authored documents and revision | Provisional registry-level hoop for the planning lead’s analytical contribution.          |
| Agent-assisted searching and critique    | Documented process claims; no inference of independent human or institutional audit.      |
| Incumbent programme operation            | Remains with its recorded operating owners and governance.                                |
| Business-case submission                 | Prepared form or retrospective assertion requires a sent artifact to authenticate filing. |
| Funding, build and customer impact       | Separate decisions and outcomes; none established by document completion.                 |

[HR/AI] The source describes the lead challenging original ideas, retaining contradictory evidence and revising the investment narrative. These are useful analytical behaviours. An internal review score, a count of searches or the number of agents is not a credential, an independently measured quality gain or evidence of programme adoption.

[AI] Evidence gaps also constrain negative claims. Unlocated action records do not prove the operating teams did nothing. A missing attribution owner does not prove no organization addressed the problem. A proposed integration is not necessarily new invention. Preserve who framed the question, who gathered evidence, who controlled execution and who could approve spending.

RELEASE CREDIT AT THE RIGHT LEVEL

[PD] Credit an authored diagnosis when its artifact and lineage are supported; credit a submitted decision only when its handoff is authenticated; credit operational improvement only with an appropriate mature comparison. Treat the current case review as internal analytical assurance, not outside endorsement.

[HR/AI] Derivative source narrative and recovered-artifact scope determine the contribution grade. No invented biography, funding, implementation, independent endorsement or realized benefit is added.

EXHIBIT 7 / NOVELTY-TO-INCREMENT CONTRACT

# Buy the evidence that changes the decision at an acceptable cost

[PD] NIEC connects scoped novelty claims to a feasible action and an incremental-value test. It preserves useful integration without granting a blank cheque to every apparent gap.

| NIEC obligation                 | Release record                                                                                   |
|---------------------------------|--------------------------------------------------------------------------------------------------|
| Define the proposed difference  | Function, original eligible population, incumbent continuation and version.                      |
| Verify equivalence and access   | Independent scope evidence, operating status, permission, capacity and unresolved sources.       |
| Keep feasible alternatives      | Reject this cycle, integrate, build or defer; no review grants authority.                        |
| Price both mistakes             | Wrong rejection, negative-value action, delay, review and displaced work.                        |
| Compare information investments | Common-evidence rule, same-packet optimizer, fuller review and a calibrated information ceiling. |

[PD] In the invented benchmark, the base audit lowers decisive errors but costs more than its default. Even the best rule using that same paid evidence cannot beat the cheaper common-evidence rule on expected priced loss. The latter has higher error risk, so it is not an automatic recommendation. A more accurate, separately priced review performs better under the assumed law; real accuracy and attainable cost remain unknown.

[PD] A larger sample cannot fix negative expected value. The corroborated screen can detect its own break-even against the weak default, yet remains insufficiently precise against the stronger common-evidence rule and is dominated by the modeled fuller review. A purchase decision needs the right comparator and an approved absolute risk criterion.

THE NEXT BOUNDED ACTION

[PD] Recover one original candidate, its incumbent comparison and the evidence needed to reverse a verdict. Calibrate review likelihoods and error costs on a blinded, original-frame sample before a shadow test. Any later operating trial requires its own approval, capacity plan and mature physical outcomes.

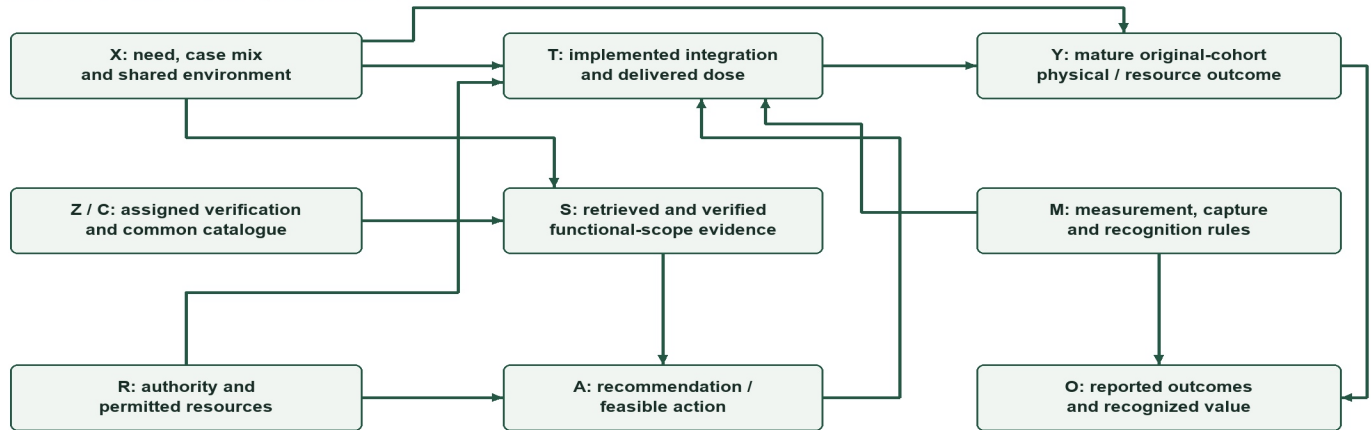
[PD] All model quantities are independent hypothetical decision units. No historical novelty, causal increment, achieved field power or operational launch is implied.

## TECHNICAL ANALYSIS

# Separate novelty, equivalence and incremental value

[PD] The Novelty-to-Increment Evidence Contract (NIEC) binds a scoped equivalence claim, the searched evidence universe, feasible action rights and a test of added value.

## NIEC: SCOPE EVIDENCE, FEASIBLE CHOICE AND ADDED VALUE



Proposed paths. Measurement M may change behavior T; shadow Z estimates decision quality, not physical benefit.

[PD] Z assigns verification; C is existing catalogue evidence; S is the resulting scope evidence; A is the feasible recommendation. Authority R constrains A and implementation T. T can change mature physical/resource outcome Y. Measurement M changes recorded O and can also change T. Context X influences retrieval, delivery and outcomes. The shadow study tests recommendations, not the downstream physical path. Conditioning on verification or survival can select on both evidence and apparent novelty.

[HR/AI] The registry describes a broad novelty pass and a later business-case screen, with different candidate frames. Keep live, in-development, rescope and missing evidence distinct. A retrieved programme name or URL does not establish equivalent action, eligible population, geography, timing, interface or authority. A catalogue omission does not establish global absence. Henderson and Clark distinguish component reuse from architectural change; an integration can add value without inventing a component.

[HR/AI] Prepared final-case language and claimed submission coexist with explicit lack of adoption evidence. Exact verdict, Pareto and business-case originals were not recovered; one related planning artifact cannot authenticate them. Repeated memory, byline and default metadata do not establish sole authorship or chronology. Contribution is a provisional registry-level hoop for scope-setting, challenge and preparation, not accepted funding, implementation or realized benefit.

Basis: proposed DAG [PD]; bounded source interpretation [HR/AI]. Henderson/Clark (1990), Carroll et al. (2007), Morwitz/Fitzsimons (2004).

TECHNICAL ANALYSIS

# A selected rejection rate cannot validate the screen

[AI] Independent construction: 100 candidates; the initial screen labels 20 as novel and selects them for verification. Assume a perfect reference solely for this proof.

[AI] Of the selected 20, twelve are equivalent duplicates and eight are genuinely novel. Here positive means initially labelled novel; reference-positive means truly novel. The selected rejection fraction is  $12/20=60\%$ , a false-discovery fraction under the assumed reference. The population false-positive rate divides false positives by all reference duplicates, including those the initial screen rejected.

| Full reference world    | TP | FP | TN | FN | FPR |
|-------------------------|----|----|----|----|-----|
| A: 16 duplicates in all | 8  | 12 | 4  | 76 | 75% |
| B: 60 duplicates in all | 8  | 12 | 48 | 32 | 20% |

[AI] Both worlds produce exactly the same reviewed result. Over all feasible duplicate totals from 12 through 92, FPR ranges from  $12/92$  to 1. Neither the population error rate nor missed useful candidates is identified by the selected review. In the historical record, the reviewer labels themselves also lack independent reference validation. Begg/Greenes explains why selective verification matters; no missingness correction is assumed here.

| Same fixed original cohort of 1,000 | Parent only failures | Parent + instrument | Added prevention |
|-------------------------------------|----------------------|---------------------|------------------|
| World zero                          | 80                   | 80                  | 0                |
| World benefit                       | 80                   | 70                  | 10               |
| World harm                          | 80                   | 90                  | -10              |

[AI] All worlds start with 100 failures without the parent and allocate half its planned reduction of 20 to the instrument: booked credit is 10 in every world. The true addition is the parent-only versus parent-plus-instrument contrast. Non-additive bookkeeping avoids one stacking convention; it cannot establish causal incrementality. These counts are invented, not rescaled source values.

[AI] Even a negative search needs detection assumptions. With prior existence probability  $1/2$ , no false hits and recall  $q$ ,  $P(\text{exists} \mid \text{no hit})=(1-q)/(2-q)$ : recall 0.8 gives  $1/6$ ; recall 0.2 gives  $4/9$ . Shared-corpus searches cannot be multiplied as independent evidence. No field prior or recall is known.

Basis: exact enumerated hypothetical proof [AI]. Begg/Greenes (1983); Altman/Bland (1995) concerns statistical negative evidence, not a search-recall theorem.

## TECHNICAL ANALYSIS

# Validate what the instrument actually observes

[PD] The measurement contract has separate records for document existence, functional equivalence, accepted ownership, implemented dose, physical outcome and recognized financial credit.

[HR/AI] Task closure is a proxy for execution. A recorded action can be late, delivered to the wrong unit, incomplete or operationally ineffective. Conversely, missing proof is not proof that no action occurred. Leading-indicator movement is not an effect estimate. A full Pareto or zero residual establishes category closure, not correct classification, physical custody or unique affected outcomes. A common maturity factor cannot validate every action-specific lag.

[PD] Freeze a candidate-level identity and version map across initial proposals, rejects, rescopings, missing records and survivors. Define equivalence using function, population, time, authority and existing operational coverage. Sample all strata with recorded inclusion probabilities; independently reconstruct reference evidence at the decision deadline. Report missing-reference bounds rather than dropping unresolved candidates. Correlated sources remain one evidence lineage.

[PD] Duplicate coders blinded to arm, initial verdict and later outcome assess equivalence and delivered action. Report weighted confusion, prevalence, raw agreement and cluster-bootstrap kappa, then resolve disagreements against independent artifacts. Four hundred independent binary assessments have worst-case agreement precision near plus/minus 4.9 percentage points; this is neither kappa power nor cluster-adjusted precision. Check invariance across region, runtime version, integration type and evidence availability.

[PD] For a later operating test, link the original eligible cohort to action content, frequency, timing, coverage and quality, following Carroll and colleagues. Preserve original customer deadlines; a later scan or recovered claim does not retroactively erase a miss. Validate delayed capture with held-out vintages and reopens, and include false alerts, review work, ordinary-flow disruption and maintenance. Morwitz/Fitzsimons shows measurement can alter behavior; its consumer experiments supply no operating effect size.

[AI] Annualizing one snapshot is a scenario, not a forecast. Benefit share, incremental cost avoidance and released cash need separate denominators and evidence. Recognition capture can change credit while physical outcomes stay constant; unknown Finance validation is neither zero true value nor proven success.

Basis: source boundaries [HR/AI] and proposed measurement [PD]. Carroll et al. (2007), Morwitz/Fitzsimons (2004), Cohen (1960).

## TECHNICAL ANALYSIS

# Compare verification with the best permitted common evidence

[PD] First test a shadow recommendation rule; test physical implementation separately only after authority, feasibility and measurement are established.

[PD] Freeze original packets and evidence rights before assignment. Randomize independent reviewer domains; retain blocked, rejected and unresolved packets. Advice locks at 72 hours for a following 90-day window; the hidden constructed reference is released 30 days afterward. Catalogue hits are common evidence. Review costs apply to every assigned packet. Compare fixed actions, the prespecified best common-hit rule, the best permitted own-packet signal rule, and full review. Other packets' signals are unavailable at lock; shared review conditions still induce dependence.

[PD] Let  $L$  be priced feasible-decision regret per original packet, including missed useful integration, unproductive build, disruption and one-cycle delay. The primary estimand is  $E[L(\text{usual}) - L(\text{NIEC})]$ ; subtract the additional assigned review price  $K$  in the same units. Safety is the unconditional frequency of wrong feasible rejection or negative action-net authorization, before review fees. Blocked packets have a restricted feasible set and are not erroneous kills for unavailable actions. Reduction relative to a weak usual rule does not establish acceptable absolute risk or superiority to cheaper evidence.

[PD] The compact benchmark uses constructed latent scope/value states and fallible evidence, not inferred historical probabilities. Exact shared-domain moments and selected finite marginal tails check the stated law. Normal power is a conditional planning approximation. Break-even must lie inside the allowed response path; otherwise report it unreachable. Do not extrapolate MDE, call selected tails uniform coverage, or interpret a large assumed full effect as evidence of field feasibility. Calibrate variance, prevalence, reference error and recruitment in an independent pilot before fixing a confirmatory design.

[PD] A later approved operating trial compares parent continuation with parent plus the integration on original eligible cohorts, retaining customer remedies. Assign at the level containing shared queues and spillovers; measure mature original-promise failures and full resources by intention to treat. Prespecify baseline version, region, task type, prior evidence quality and available capacity as modifiers. Do not condition total effects on realized task closure, captured flags or post-treatment survival.

Basis: prospective protocol and estimands [PD]. Hudgens/Halloran (2008); Carroll et al. (2007). All modeled values are synthetic; source records supply no calibrated law.

TECHNICAL ANALYSIS

# Prespecify what would reject a claim or hold a decision

[PD] Falsifiable propositions distinguish evidence against a mechanism from a funding gate that merely lacks precision.

| Proposition                                               | Prespecified rejection or hold rule                                                                                                                                  |
|-----------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| P1: the selected rejection rate identifies population FPR | Reject without the full reference-negative denominator and validated reference; the two-world construction disproves identification from selected outcomes alone.    |
| P2: an existing component makes integration redundant     | Reject when matched-scope parent-plus-integration shows material added benefit after full cost; otherwise hold until equivalence and causal comparison are observed. |
| P3: task closure is adequate proof of delivered dose      | Reject at a prespecified material discordance threshold against blinded content/timing/coverage evidence; retain missing proof as unknown.                           |
| P4: allocated parent credit identifies added benefit      | Reject as an identification claim; the same credit admits positive, zero and adverse increments. Test the added intervention directly.                               |
| P5: paid verification improves the feasible decision      | Require economic clearance and a justified absolute error guardrail against the strongest common-evidence comparator; imprecision is a hold, not proof of no effect. |

[HR/RC] The private inventory contains 38 diagnostic vectors: one weak ordered outcome series, 14 version comparisons, 12 design trajectories and 11 profiles. The weak series has no fitted X or moving-range crossing; it does not establish a stable phase. Tentative XmR uses 2.66 and 3.267. No eligible Laney panel exists and none is computed. Missing weekly points and action-specific lag curves are not invented.

[PD] Under an assumed linear conditional-mean model, expected untreated follow-up after selection at baseline  $x$  is  $m + \rho \times (x - m)$ , where  $m$  is the population mean and  $\rho$  is baseline persistence. This illustrates regression to the mean, not an estimated law here. For a later observational rollout, pretrend plots, equivalence bounds, negative controls and sensitivity analyses assess assumptions; nonsignificance does not prove parallel trends or absence of effect.

### Limitations and identifying assumptions

[AI/PD] Exact historical verdicts, production exposure, joined customer outcomes and accepted funding are unauthenticated. Search coverage and equivalence-reference error are unknown; no archived candidate census is invented. The shadow law assumes its finite value support, shared environment, evidence rights and prices. It cannot reveal real counterfactual integration value. Deployment needs independent calibration of timing, capacity, interference, reference maturity and adverse workload; recurrent deferral must be repriced beyond this cycle.

Basis: bounded source/SPC assessment [HR/RC], new identification results [AI], testable evaluation plan [PD]. Altman/Bland (1995), Begg/Greenes (1983), Barnett et al. (2005).

PROPOSED DECISION TEST / 1 OF 4

# Separate an incumbent match from a useful eligible gap

[PD] A synthetic shadow review chooses reject, integrate, build or defer. It changes advice only. Each original proposal has one hidden constructed state, not a field-observable causal label.

| State / prior          | Function F, gap G | Integrate net | Build net |
|------------------------|-------------------|---------------|-----------|
| Duplicate / 0.30       | 1, 0              | -2            | -5        |
| Integration gap / 0.30 | 1, 1              | 5             | 2         |
| New function / 0.20    | 0, 1              | -2            | 7         |
| Unviable / 0.20        | 0, 0              | -3            | -6        |

[PD] Integrate gross is (0,7,0,0), work 2; build gross (0,7,12,0), work 5. Either action in unviable adds disruption 1. Reject yields 0; defer costs 1 for this cycle. Common catalogue-hit probabilities are (0.90,0.80,0.15,0.35). Usual rejects on a hit and builds otherwise.

| Evidence: clear / fragile | Capture     | F accuracy  | G accuracy  |
|---------------------------|-------------|-------------|-------------|
| Base                      | 0.90 / 0.60 | 0.85 / 0.65 | 0.80 / 0.60 |
| Stress                    | 0.75 / 0.40 | 0.85 / 0.65 | 0.65 / 0.50 |
| Corroborated              | 0.95 / 0.80 | 0.95 / 0.85 | 0.95 / 0.85 |

[PD] Three packets share hidden grade, clear with probability 0.65; domains are independent. Conditional on state/grade, captured flag errors and catalogue evidence are independent. Audit rejects on G=0, integrates on G=F=1, builds on G=1/F=0 and defers if capture fails. Flags are fallible; policy never sees truth.

AUTHORITY IS A HARD CONSTRAINT

[PD] Known 10% blockers permit reject/defer only; reject has zero feasible regret. All review fees remain paid. Other packets' signals are unavailable. Original advice locks at 72 hours; a following 90-day window has hidden constructed reference released 30 days later. No real auditor observes every alternative effect.

[PD] Entirely invented. Rejection means no funding this cycle, not global worthlessness. Common intake and blinded research-reference costs are separately budgeted. Independent reference and transport calibration precede any field use.

# No sample size repairs an unreachable break-even

[PD] Regret is best feasible state-specific net minus chosen net. Gates require lower regret by more than the review fee K and fewer decisive errors on all original packets.

| Domains / arm      | 80% MDE      | Fee K | Zero-test power at BE | Full-law joint floor |
|--------------------|--------------|-------|-----------------------|----------------------|
| Small / 16         | not attained | 0.50  | unreachable           | 0.0%                 |
| Base / 128         | not attained | 0.50  | unreachable           | 0.8%                 |
| Stress / 128       | not attained | 0.50  | unreachable           | 0.0%                 |
| Corroborated / 256 | 0.376        | 0.90  | >99.9%                | 98.7%                |

[PD] A decisive error is wrong rejection of useful integration/new work, or authorization of negative action net before review fees. Blocked rejection is not an error. Defer has zero decisive-error risk but delay and foregone value. Positive-value misallocation remains priced in regret.

[PD] Audit is used with probability lambda per original packet, usual otherwise; K is fully paid. Exact shared-grade moments set fixed normal radii using the largest variance on lambda in [0,1], with nominal directional error 2.5%. Base maximum gain 0.40779 is below K=0.50; stress gain is negative. Neither has a break-even on this path.

[PD] Corroborated MDE 0.37592 is below K=0.90. Its normal joint floor is 98.73%; exact marginal tails imply a 98.695% joint lower bound. Exact loss-null tails are 1.97037% base and 2.50161% corroborated; risk-null tails 2.47720%/2.47189%. The corroborated economic-boundary false clearance is 2.08502%. The slight nominal excess is retained.

### TEST THE STRONGER COMPARISON

[PD] Corroborated net gain versus the common-evidence rule is only 0.18882. At 256 domains/arm its normal lower bound at expected means is -0.03697 and economic-clearance power 37.4%. Fuller review has lower priced loss and risk. The 1,536-packet recruitment is hypothetical, not a feasible or approved launch.

[PD] Four direct allocations; 18 selected exact marginal tails; no Monte Carlo or minimum search. Base midpoint is a diagnostic, not a break-even boundary. Fixed normal coverage is conditional, not uniform. Zero joint floor is uninformative, not zero joint power.

PROPOSED DECISION TEST / 3 OF 4

# Preserve both error types and every actual review cost

[PD] Entries show total regret including review fees / decisive-error risk defined before fees. Units are hypothetical resources, not dollars or customer-harm rates.

| Policy                | Base            | Stress          | Corroborated    |
|-----------------------|-----------------|-----------------|-----------------|
| Catalogue / usual     | 2.2680 / 38.70% | 2.2680 / 38.70% | 2.2680 / 38.70% |
| Always reject         | 2.6100 / 45.00% | 2.6100 / 45.00% | 2.6100 / 45.00% |
| Always integrate      | 2.7000 / 63.00% | 2.7000 / 63.00% | 2.7000 / 63.00% |
| Always build          | 3.2400 / 45.00% | 3.2400 / 45.00% | 3.2400 / 45.00% |
| Always defer          | 3.6100 / 0.00%  | 3.6100 / 0.00%  | 3.6100 / 0.00%  |
| Best common evidence  | 1.9170 / 47.70% | 1.9170 / 47.70% | 1.9170 / 47.70% |
| Proposed audit        | 2.3602 / 20.15% | 3.0139 / 22.95% | 1.7282 / 7.73%  |
| Best same-packet rule | 1.9621 / 29.47% | 2.3262 / 44.46% | 1.5649 / 12.62% |
| Fuller review         | 1.3680 / 3.06%  | 1.3680 / 3.06%  | 1.3680 / 3.06%  |

[PD] Fees are 0.50 base/stress, 0.90 corroborated and 1.20 fuller review, including blockers/unknowns. Fuller capture and each flag accuracy are 0.98. Common evidence has 16 mappings; all hit-by-scope mappings are optimized in two separable 1,024-rule blocks. Best expected loss is not an approved risk-constrained optimum.

[PD] Base net versus usual =  $a \cdot h \cdot 1.5574095 - 1.1496150 - 0.50$ . Favorable state-by-resource-category changes alone receive realization  $a$  and nonduplicated share  $h$ ; all added burden, work and fees remain fully charged. Within-action overlap is already removed from constructed gross value.

| Net units / packet | h = 0.50 | h = 0.75 | h = 1.00 |
|--------------------|----------|----------|----------|
| a = 0.25           | -1.4549  | -1.3576  | -1.2603  |
| a = 0.50           | -1.2603  | -1.0656  | -0.8709  |
| a = 1.00           | -0.8709  | -0.4816  | -0.0922  |

[PD] Regret includes work, disruption, delay and foregone value once. Action net excludes review fee; total loss includes it. Any real duplicate adjustment needs common-unit evidence; a shared parent or defect pool is not sufficient.

# Use an information ceiling before commissioning more review

[PD] Perfect state information can remove common-evidence regret 1.917. For the base evidence actually offered, the maximum priced-loss improvement before its fee is only 0.454923, below the 0.50 fee.

[PD] This ceiling ranges over all same-packet rules and holds only under the locked prior, support and expected-loss objective. It does not rule out paying for a separately approved risk target. At build price 3 rather than 5, base audit costs 2.40498 versus common evidence 1.575; fuller review costs 1.35766. Cost calibration can change the preferred alternative.

[PD] A separate research choice has states: equivalence/access evidence useful (0.40), incremental-response evidence useful (0.35), neither (0.25). Payoffs for those two studies and deferral are (6,-2,-2), (-1,7,-3), (0,0,0), including downstream costs. Signal accuracy q has symmetric errors; enumerate all 27 rules.

| Research scenario | q     | EVSI    | Cost    | Net EVSI |
|-------------------|-------|---------|---------|----------|
| Low               | 50.0% | 0.71250 | 0.50000 | 0.21250  |
| Base              | 75.0% | 2.04375 | 1.00000 | 1.04375  |
| High              | 90.0% | 2.94750 | 2.00000 | 0.94750  |

[PD] No-information value is 1.30; EVPI is 3.55. Higher accuracy need not maximize net information value. These are alternative hypothetical likelihoods and prices, not confidence intervals. Research value is not added to operating benefit or to the shadow-review gain.

VALIDATE THE REFERENCE AS WELL AS THE POLICY

[PD] At main prices, regret support is 0..9. Reference reassignment within that support on epsilon of original packets per arm permits 18\*epsilon loss and 2\*epsilon risk-contrast error, verified fees fixed. Changed prices or support need a new bound. Preserve unavailable sources and benchmark authority; calibrate before any operating experiment.

[PD] The 72-hour lock and future-window reference cannot manufacture causal truth. A later field trial requires authorized actions, original customer eligibility, mature physical outcomes and full interference-aware costs. Shadow advice improvements establish none of those effects.

REPRODUCIBILITY / READER PROTOCOL

# Reproduce the search boundary and the incremental claim

[AI] The Novelty-to-Increment Evidence Contract makes the match, the decision and the outcome separately challengeable.

| Reproduction layer      | Required record and falsification test                                                                                                                                                                                          |
|-------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Candidate frame         | Freeze the proposal, origin, version, screening stage and eligibility. Preserve rejected and unresolved candidates. Different passes cannot form a cohort funnel without candidate-level linkage.                               |
| Existing capability     | Record the incumbent function, served cohort, geography, operating status and execution rights. Recover dated evidence and owner confirmation. A shared label or technical component does not establish equivalence.            |
| Verification quality    | Maintain search coverage and missing-reference reasons. Independently assess samples from every verdict class. Report class-specific errors and coverage; rejection yield is not a population false-positive rate.              |
| Increment and economics | Define continuation with the incumbent alone, the marginal action package, mature outcome and avoidable resources. Allocation of an existing benefit is an accounting convention; it cannot identify the enabler effect.        |
| Decision and delivery   | Preserve the available evidence, authorized alternatives and decision lock. Link acceptance to staffing, implementation and mature results. Authorship, a verification report and a proposed pilot do not establish that chain. |

[PD] Repeat the economic comparison under plausible false-match and missed-match rates. Charge review cost to unresolved cases. Include useful integration, genuinely new functionality, rejection and a time-bounded hold. Preserve the option with the greatest supported value, rather than maximizing a rejection quota or the number of surviving ideas.

[AI] The private archive retains derivative source records, recovery limits, version forks, executable calculations and internal review. It does not authenticate an unrecovered original. Public numerical examples are independent constructions; they are not rescaled historical metrics.

Protected registry and bounded recovery record; proposed NIEC reproduction protocol. Original identities, raw quantities and internal source locations remain private.

## REFERENCES / 1

# Primary sources and their role in this case

[AI] These sources inform verification, implementation and evaluation. They do not certify the historical novelty verdicts or financial estimates.

NIST/SEMATECH. Individuals Control Charts, e-Handbook section 6.3.2.2. Calculate tentative limits only for comparable ordered observations. A short series with unknown phase cannot establish a stable process; catalogue versions and proposal counts are not a process series.

Laney, D. B. (2002). Improved Control Charts for Attributes. *Quality Engineering*, 14(4), 531–537. Adjusted rate charts require actual count-exposure subgroups with coherent definitions. Neither one snapshot nor rounded rates supply the missing panel.

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20, 37–46. Code existence, functional equivalence, implementation status and execution rights separately. Independent agreement does not establish completeness or correctness of the search.

Barnett, A. G., van der Pols, J. C., & Dobson, A. J. (2005). Regression to the Mean: What It Is and How to Deal with It. *International Journal of Epidemiology*, 34, 215–220. Worst-unit action selection can combine temporary extremes with persistent conditions. Freeze eligibility and compare unselected units before assigning subsequent convergence to instrumentation.

Callaway, B., & Sant'Anna, P. H. C. (2021). Difference-in-Differences with Multiple Time Periods. *Journal of Econometrics*, 225, 200–230. For later phased implementation, distinguish access to evidence from completed corrective action. Estimate group-time effects with coherent timing, mature outcomes and an explicit untreated comparison.

Collier, D. (2011). Understanding Process Tracing. *PS: Political Science & Politics*, 44, 823–830. Grade analytical direction separately from search execution, submission, funding and operations. Repeated derivative narrative is not independent corroboration.

Strong, M., Oakley, J. E., Brennan, A., & Breeze, P. (2015). Estimating the Expected Value of Sample Information Using the Probabilistic Sensitivity Analysis Sample. *Medical Decision Making*, 35, 570–583. Value additional verification by changes in the best feasible decision. Price unresolved reviews, wrong rejection and false funding; compare with strong fixed and common-evidence rules.

Primary author, journal, institutional and official methods sources. Access scope and case-specific application are recorded privately.

## REFERENCES / 2

# Primary sources and their role in this case

[AI] These sources inform verification, implementation and evaluation. They do not certify the historical novelty verdicts or financial estimates.

The W. Edwards Deming Institute. The Funnel Experiment. A sequence of small metric changes does not justify repeated operational adjustment. Protect an interpretable process definition before changing the action policy.

Campbell, D. T. (1976). Assessing the Impact of Planned Social Change. A rewarded novelty or rejection score can alter proposal wording and search behavior. Preserve rejected candidates and independently assess both classes.

Goodhart, C. A. E. (1992). The Objectives for, and Conduct of, Monetary Policy in the 1990s. Reserve Bank of Australia Annual Conference, section 1(b)(i). A consequential target can weaken its relationship with the intended outcome. Proof completion, search volume and kill counts need independent validation against decision quality.

Ittner, C. D., & Larcker, D. F. (2003). Coming Up Short on Nonfinancial Performance Measurement. Harvard Business Review, 81(11), 88–95, 139. Test the chain from information to implemented action and mature customer outcome. An attribution share or a completed task record cannot validate that chain.

Tucker, A. L., & Edmondson, A. C. (2003). Why Hospitals Don't Learn from Failures: Organizational and Psychological Dynamics That Inhibit System Change. California Management Review, 45(2), 55–72. An unresolved evidence gap should trigger correction of the operating mechanism, not a recurring narrative workaround. Assign evidence ownership and escalation authority.

OECD / European Commission JRC (2008). Handbook on Constructing Composite Indicators: Methodology and User Guide. Weights and aggregation rules embed judgments. A composite review score does not establish search recall, causal increment or a fundable opportunity.

Henderson, R. M., and Clark, K. B. (1990). Architectural Innovation: The Reconfiguration of Existing Product Technologies and the Failure of Established Firms. Administrative Science Quarterly, 35(1), 9-30. doi:10.2307/2393549. Existing components do not establish equivalent architecture, integration or operational function.

Primary author, journal, institutional and official methods sources. Access scope and case-specific application are recorded privately.

## REFERENCES / 3

# Primary sources and their role in this case

[AI] These sources inform verification, implementation and evaluation. They do not certify the historical novelty verdicts or financial estimates.

Begg, C. B., and Greenes, R. A. (1983). Assessment of Diagnostic Tests When Disease Verification is Subject to Selection Bias. *Biometrics*, 39(1), 207-215. doi:10.2307/2530820. Audit rejected and unverified proposals as well as survivors; selected verification cannot identify population screening accuracy without assumptions.

Altman, D. G., and Bland, J. M. (1995). Statistics notes: Absence of evidence is not evidence of absence. *BMJ*, 311(7003), 485. doi:10.1136/bmj.311.7003.485. Distinguish a bounded negative result from evidence that a material effect or existing mechanism is absent.

Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., and Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science*, 2, 40. doi:10.1186/1748-5908-2-40. Measure delivered content, dose, timing, coverage and quality separately from task closure and outcome effectiveness.

Morwitz, V. G., and Fitzsimons, G. J. (2004). The Mere-Measurement Effect: Why Does Measuring Intentions Change Actual Behavior? *Journal of Consumer Psychology*, 14(1-2), 64-74. doi:10.1207/s15327663jcp1401&2\_8. Treat instrumentation as potentially behavior-changing and include its effort and effects in the operating comparison.

Naaktgeboren, C. A., et al. (2016). Anticipating Missing Reference Standard Data When Planning Diagnostic Accuracy Studies. *BMJ*, 352, i402. Plan reference verification for the candidate frame, including rejected and unresolved cases. Selected-only accuracy can be biased; corrections require assumptions about missingness. Applied by analogy to proposal screening, not as validation of any historical search.

Hudgens, M. G., & Halloran, M. E. (2008). Toward Causal Inference with Interference. *Journal of the American Statistical Association*, 103, 832-842. Original manuscript. Contain shared reviewer, programme and operational spillovers within assignment domains or model them explicitly.

Primary author, journal, institutional and official methods sources. Access scope and case-specific application are recorded privately.

# Preserve useful integration while challenging false novelty

[AI] Author-led, agent-assisted review; not external peer review. The case evaluates what a novelty screen can establish and what requires an incremental test.

| Major issue                                  | Resolution and decision boundary                                                                                                                                                              |
|----------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| A rejection fraction called an error rate    | Define the classifier, reference truth and full eligible frame. The same selected audit is compatible with different population error rates. Historical search accuracy remains unidentified. |
| A catalogue match treated as duplication     | Compare function, cohort, deployment state and action rights. Existing technology can support a useful integration. A missing match does not prove universal absence or commercial value.     |
| Search verdicts treated as independent truth | Distinguish reported screening from authenticated equivalence. Search intensity and agent count do not establish calibration, independence or complete coverage.                              |
| Missing proof treated as failed execution    | Separate action completion, evidence capture, linkage and mature effectiveness. A zero linked-proof observation cannot identify which mechanism failed.                                       |
| A smaller benefit share treated as increment | Keep accounting credit separate from the parent-only counterfactual. The same allocated share is compatible with positive, zero or harmful incremental effect.                                |
| Conservative rejection treated as costless   | Price valuable ideas incorrectly killed as well as duplicate work funded. Compare the full decision loss, unresolved workload and verification fee with feasible alternatives.                |

## Partner critique: can a COO act on page one?

[AI] Yes. Require a typed comparison with the incumbent and a documented decision owner. Keep useful integration eligible, subject its marginal benefit to testing, and buy additional verification only when it can change a feasible decision enough to cover its cost. A high rejection rate is neither the goal nor the investment case.

## Remaining limits and minor issues

[AI] The technical layer consolidates source authentication, submission status, missing reference truth, measurement, transport, attribution and economics limits. Neither the search-error examples nor the shadow decision model estimates historical impact. Internal review closes publication issues, not those empirical questions.

### REVIEW DISPOSITION / PASS WITHIN STATED SCOPE

[AI] Publish the bounded evidence reconciliation, constructive counterexamples and proposed evaluation. A reproducible screening policy supports further testing; it does not authorize funding or establish a delivered operating result.

Final source, independent numerical, methods, partner and actual-page reviews required. Internal publication review is distinct from external assurance.