

Benefit Realization & Incrementality

A reporting recovery is not a new programme benefit

An operating-project case on realization exposure and the incrementality test for a carryover portfolio. It separates missing actuals, future forecasts and expansion value from the value of maintaining an existing programme.

REALIZATION & INCREMENTAL VALUE

BLUF / BOTTOM LINE UP FRONT

Recover the missing actuals. Test the claim against a defined continuation case.

ANALYTICAL FINDING

The source combines recognized period amounts with future forecasts. Better reporting can increase recorded realization without changing delivery; even complete actuals do not identify the benefit of expansion over continuation.

CONTRIBUTION IN THE SOURCE RECORD

Registry-level hoop: reported analysis, disclosure and escalation; no ownership of Finance approval or programme outcomes.

EVIDENCE STATUS

Evidence-reconciliation and proposed-design case. Originals and Finance closure remain unauthenticated; implementation and impact are unestablished.

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

STRUCTURED ABSTRACT

Decide what is being recognized before deciding what is incremental

[AI] A recorded shortfall can combine incomplete reporting, weak delivery and an unsuitable comparison. These require different evidence and different decisions.

Element	Decision-relevant conclusion
Context [HR]	A planning lead assembled Finance quotations, reported realization arithmetic and prepared a disclosure-and-escalation response for carryover claims. The sizing ruling and incrementality response are not established.
Objective [AI]	Separate past recognized amounts, future continuation, proposed expansion and the authority to claim each. Preserve uncertainty without treating maintenance as worthless.
Approach [RC/AI]	Type each metric by period, status and reporting level; reconcile cohort and booking clocks; test whether the stated baseline is a genuine counterfactual.
Findings [AI]	A year-to-date recognition ratio is not a full-year forecast ratio. A broader reporting line is not one project. Recovering reports does not create a physical increment.
Decision [PD]	Finance confirms recognition and the comparison basis. The sponsor records the sizing ruling. Each programme owner supplies scoped actuals and a separately identified continuation-versus-expansion case.
Implication [PD]	Use the proposed Realization-to-Incrementality Evidence Contract (RIEC). Calibrate a shadow reporting test before operational use; the hypothetical richer rule loses to a cheaper rule using the same current evidence.

[PD] Release an incremental claim only with a common unit, cohort, horizon, continuation comparator and accepted authority. Unsized is a visible decision state with an owner and deadline, not a declaration of zero economic value.

[HR] historical record; [RC] recalculated; [AI] analytical inference; [PD] proposed design. Protected source S1. Original identities, figures, dates and locators remain private.

EXHIBIT 1 / METRIC IDENTITY

Do not rank a realized period amount against a future forecast

[AI] Similar-looking actual-to-plan fractions can describe different objects. Division does not make their periods, populations or evidentiary status comparable.

Reported object	Required denominator	Permitted interpretation
Recognized period amount	Same-period, same-scope plan.	Recognition against that plan, subject to reporting completeness.
Full-year outlook	Full-year commitment on the same basis.	A forecast of attainment, not a completed-year result.
Broader reporting line	Its own portfolio commitment.	Context for a nested project; not project-specific performance.
Parent portfolio total	Same parent scope and period.	Portfolio context; no automatic allocation of credit to one workstream.

[HR/AI] The source places a period recognition ratio beside full-year forecasts and a broader roll-up. The line containing a project can be relevant to review risk without becoming that project's own performance measure. Parent favourability cannot validate a weak child claim either.

TYPE THE NUMBER BEFORE USING IT

[PD] Add period, recognition/forecast status, scope, gross/net basis, valuation rule and source vintage to every amount. Reconcile incompatible extracts. Do not average, pool or rank the ratios until those definitions agree.

[RC/AI] The private arithmetic reproduces the printed divisions and gaps while preserving their distinct meanings. No historical rates or indexed coordinates appear here.

EXHIBIT 2 / REPORTING RECOVERY

Missing reports leave delivery uncertain in both directions

[AI] A reporting failure can conceal real value, coexist with poor execution, or reflect more than one problem. The missing records do not choose among those explanations.

Independent invented fixture	Before backfill	After backfill
Same mature physical amount	100 units	100 units
Recognized amount from that cohort	30 units	100 units
New physical increment caused by backfill	0 units	0 units

[PD] The recorded increase of 70 in this fixture is a change in ascertainment, not a new operating benefit. It may matter to recognition or planning, but those decision outcomes need their own evidence and cannot be added to the same programme benefit twice.

[HR/AI] The historical extracts describe missing actuals and some entitlement being yielded; they do not quantify what complete reporting would have shown. A reporting-only diagnosis is therefore too strong, as is treating low recorded realization as proof that the programme failed physically.

RECOVER, THEN RECONCILE

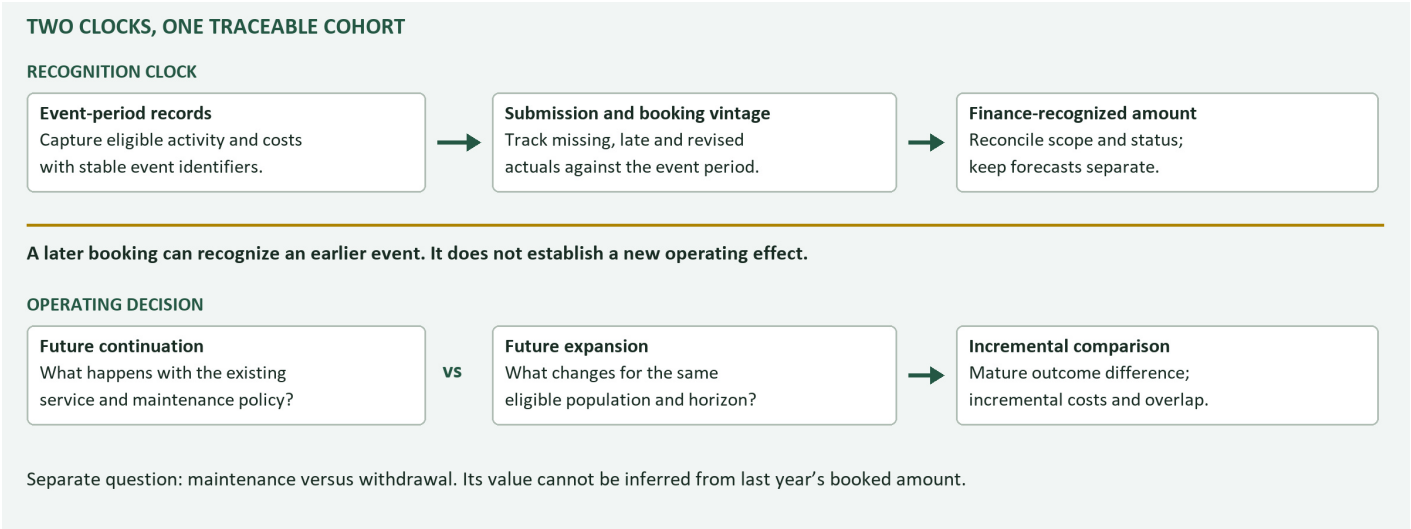
[PD] Assign a monthly actuals owner, but also verify event identity, scope, prior bookings and reversals. Preserve the original activity cohort when late reports arrive. Measure submission completion separately from independent operating outcomes and accepted recognition.

[PD] All quantities are invented. The fixture holds physical outcomes fixed by construction and illustrates a possible reporting-only change, not the historical counterfactual.

EXHIBIT 3 / TWO CLOCKS

Bridge the event cohort to the booking vintage without inventing new value

[AI] Work happens on one clock; evidence and recognition arrive on another. A late entry must remain linked to its original cohort and comparison.



[PD] Reconcile the operational ledger to the recognition ledger by original event, programme, period, valuation and accepted version. Identify first recognition, backfill, correction and reversal separately. A change in the accounting load date must not reset the benefit clock.

[AI] This bridge can show where amounts moved and which records are missing. It cannot establish that expansion changed the outcome relative to continuing the existing programme. That question belongs to a separate causal comparison.

MAKE THE CADENCE AUDITABLE

[PD] The programme owner supplies actuals; Finance accepts or rejects recognition on an explicit basis. The portfolio sponsor tracks unanswered questions and sizing decisions in the decision register. Capture exceptions, rework and overdue responses instead of counting a named owner as an operating control.

[PD] Conceptual ledger diagram with no historical quantities. Timely reporting, accepted recognition and identified incrementality are separate outcomes.

EXHIBIT 4 / CONTINUATION COMPARATOR

The same complete actuals can support positive, zero or negative incrementality

[AI] The expansion question is what changed relative to the same future period without the expansion. The prior booked amount is not automatically that comparator.

Independent constructed world	Future actual amount	Continuation amount	Increment
Expansion adds value	100	60	40
Expansion adds nothing	100	100	0
Expansion is adverse	100	120	-20

[PD] All three worlds can share a reported amount of 30 before evidence recovery and a complete future amount of 100 afterward. Complete reporting does not distinguish their increments. The values are controlled potential-outcome fixtures, not estimates from the project.

[HR/AI] The source asks whether new-year entitlement is already included in an existing run-rate and records an unquantified expansion answer. A baseline/delta column makes that question visible; merely filling it does not identify the no-expansion outcome.

SPECIFY THE COMPARISON BEFORE SIZING

[PD] Define the maintained programme, expansion version, eligible exposure and mature horizon. Use a defensible comparison design or report bounds under stated assumptions. Keep an unobserved counterfactual distinct from a missing administrative field.

[PD] Independently checked hypothetical construction. A causal increment can be negative even when the programme has a positive gross amount.

EXHIBIT 5 / MAINTENANCE AND EXPANSION

Preserving an existing benefit and adding a new benefit answer different questions

[AI] Carryover is not automatically worthless. Maintaining service may prevent deterioration even when expansion adds little or nothing.

Invented future option	Gross outcome value	Relevant contrast
Withdraw existing programme	30 units	Reference for a maintenance decision.
Continue existing programme	70 units	Maintenance versus withdrawal: +40.
Continue and expand	80 units	Expansion versus continuation: +10.

[PD] In this fixture, expansion versus withdrawal is +50, composed of maintenance +40 and expansion +10. Counting +50 and adding the maintenance +40 again would double credit the same contrast. Costs, displaced work and overlap still need to be deducted on the relevant option basis.

[HR/AI] The record describes a carryover mechanism with a defensible operating rationale alongside weak reported realization. Those can coexist. Neither existing deployment nor a maintenance label proves that the new claim is already recognized, incremental, or valueless.

WRITE TWO DECISIONS, NOT ONE LABEL

[PD] If maintenance is required, describe the loss avoided versus withdrawal and its full ongoing cost. Size expansion against continuing that maintained programme. Finance then determines recognition under its rules; the programme comparison does not itself grant approval.

[PD] Independent invented option values, not dollars or historical programme estimates. No actual withdrawal or service reduction is proposed by this case.

EXHIBIT 6 / DECISION CLOSURE

A send-ready question needs an accepted owner and a recorded ruling

[AI] Writing the disclosure and addressing the escalation are useful outputs. They do not establish that the question was sent, answered or used to change the submitted figure.

Obligation	Responsible role	Evidence that closes it
Missing or late actuals	Programme reporting owner	Linked cohort records, submission log and reconciled revisions.
Recognition basis	Finance authority	Explicit accepted scope, amount, period and valuation.
Continuation comparator	Programme analyst with decision sponsor	Prespecified comparison, sensitivity and independent evidence.
Claim sizing	Portfolio sponsor	Versioned ruling: retain, reduce or remain unsized; reasons and reopening trigger.
Capacity for the cadence	Resource owner	Accepted recurring workload and backup; not a name added to a column.

[HR/AI] A consequential sizing ask appeared in handoff prose but was not given its own row in the formal leadership decision list. The proposed monthly owner/cadence also lacks evidence of sustained operation. These are closure and execution gaps, not grounds to invent the later outcome.

BOUND THE PERSONAL CONTRIBUTION

[HR/AI] The planning lead is credited with reported exposure analysis, disclosure design and escalation, provisionally at registry-level hoop strength. Finance supplied the quoted measurements and held recognition authority; the sponsor controlled sizing. Programme results and approval remain outside the attributable contribution.

[HR] Historical originals and later decision closure remain unauthenticated in the reviewed corpus. Search or access limits do not establish that no answer exists elsewhere.

EXHIBIT 7 / DECISION CONTRACT

A more complete incrementality report must still justify its added cost and risk

[PD] The proposed Realization-to-Incrementality Evidence Contract preserves the measurement bridge and separately tests the comparison used for a new claim.

Gate	Required decision evidence
Reconcile realization	Common event cohort, booking vintage, recognition rule and current actuals.
Define the increment	Expansion versus a stated same-horizon continuation policy; maintenance tested separately.
Price the reporting choice	Explicit overclaim, underclaim, unsized and review costs; retain inexpensive fixed rules and full review.
Protect against false credit	An unconditional risk gate across all original dossiers, including unresolved and adverse cases.
Release under authority	Finance confirms recognition; the sponsor records sizing. A shadow test changes neither authority nor customer service.

[PD] In the hypothetical benchmark, the base replay can detect economic break-even yet rarely clears the separate false-credit gate. Costed corroboration improves precision, but a simple rule using the common current evidence has lower loss and no added false credit under the declared law. Its advantage depends on that law; more precision against a weaker comparator does not settle purchase.

NEXT ACTION

[PD] Reconcile a bounded original-cohort sample, obtain the continuation assumptions and register the unresolved sizing decision. Calibrate evidence error, review effort and the cost of an unsized response before committing to a larger replay. Do not book a reporting recovery as new operating savings.

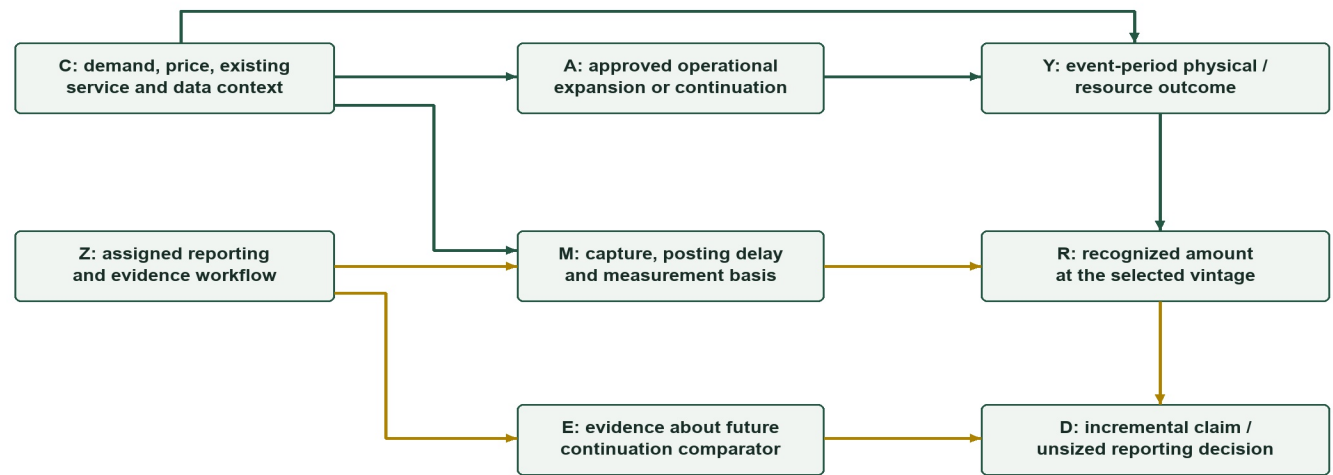
[PD] The prospective benchmark is wholly synthetic. No adopted control, Finance ruling, incremental programme effect or realized savings is claimed.

TECHNICAL ANALYSIS

Separate recognition from incremental value

[PD] The proposed Realization-to-Incrementality Evidence Contract (RIEC) links event cohorts to booking vintages, distinguishes forecasts from actuals, names the continuation alternative, and retains unknown value until the evidence supports a decision.

RIEC: OPERATING OUTCOMES AND RECOGNITION HAVE DIFFERENT PATHS



Reporting can change R and D with no change in Y. A Finance decision does not reveal both potential outcomes.

C also informs E; this baseline link is stated here to keep the drawing legible. All proposed arrows need validation.

[PD] A is the operational strategy and Y its physical or resource outcome. C captures baseline demand, prices, existing service and information quality. Z is the reporting workflow; M determines capture, delay and measurement basis; R is the recognized amount. E informs the future continuation comparator; D is the incremental-claim decision. Z can change R without changing Y. The diagram does not identify an operational effect from a reporting change.

[HR/AI] The register reports missing monthly actuals, low recognized realization, forecast shortfalls, a request to separate baseline from added entitlement, and an unresolved sizing escalation. These support a disclosure-and-evidence case. They do not establish that the programme succeeded, failed, or would have produced the same value without new spending. A reported Finance explanation is not an independent physical-outcome audit.

[AI] Keep three contrasts separate: maintaining an existing service versus withdrawing it; expanding versus maintaining that service; and improving reporting versus usual reporting. Continuing value can require recurring resources. The amount already booked in a previous year is not the future continuation counterfactual. Public-sector evaluation guidance is used as a methodological analogy, not a compliance claim.

Basis: proposed RIEC DAG [PD]; register-mediated history [HR]; Hernán and Robins (2016), HM Treasury (2026). No historical numerical coordinates are public.

TECHNICAL ANALYSIS

The same reported amount permits opposite effects

[AI] Independent hypothetical construction: every world has future gross value 100 and recognized value 30 on the same population, horizon and price basis. Only the unobserved continuation outcome differs.

Quantity in value units	World A	World B	World C
Future gross value with expansion	100	100	100
Recognized at current vintage	30	30	30
Future value under continuation	60	100	120
Added value before added cost	40	0	-20
Net after added cost 5	35	-5	-25

[AI] Even perfect recovery of the expansion amount would not select a world: the missing quantity is the contemporaneous alternative. The negative-effect world is logically admissible without an assumption that expansion cannot displace useful activity. It is not an allegation that the historical programme caused harm. A previous-year plan or booked baseline supplies neither the continuation value nor that monotonicity assumption.

[AI] A second illustration holds underlying gross value at 80 while capture rises from one quarter to three quarters. Recognition rises from 20 to 60, yet underlying value changes by zero. This assumes one fixed event cohort, no false-positive bookings and unchanged valuation. Improved submission can recover evidence of old activity; its recognition gain must not be credited again as new operating benefit.

[AI] With recognition $R=qY$, $R=20$ and justified capture q in $[0.25,0.50]$, Y lies in $[40,80]$. If continuation C is in $[30,70]$ and added cost is 5, net increment $Y-C-5$ lies sharply in $[-35,45]$ when all combinations are admissible. Both endpoints and every intermediate value can occur. These are assumption-defined support bounds, not confidence intervals. Without a positive lower capture bound or a bounded continuation outcome, this finite interval is unavailable.

[PD] ASK WHICH UNKNOWN WOULD CHANGE THE DECISION

Recovering current actuals and estimating continuation solve different problems. Choose evidence that narrows the decision-relevant uncertainty.

Basis: independent exact constructions [AI]; no historical calibration. Richardson et al. (2014) distinguishes partial identification from sampling uncertainty.

TECHNICAL ANALYSIS

Reconcile event time, posting time and scope

[PD] Keep the benefit event, operational exposure, valuation, accounting recognition and decision authority as linked but distinct records.

Field or check	Decision meaning
Original population and event period	Freeze eligible activity before observing whether an owner submitted actuals.
Posting vintage and revision reason	Track first submission, late backfill, correction, reversal and restatement against the original cohort.
Value type and price basis	Separate plan, forecast, gross avoided cost, capacity release, net value and verified cash.
Continuation and added resources	Specify what persists with existing maintenance and what the proposed expansion changes.
Accepted ownership and ruling	Record feasible monthly effort, named acceptance, received evidence and exact Finance scope.

[HR/AI] The source mixes year-to-date recognition with full-year forecasts, an unrelated precedent and a broader reporting line containing multiple programmes. These cannot form a common realized-performance league table. A favourable parent total does not prove a child contribution; absence from selected named drivers does not prove no contribution. Conflicting validation labels and unrecorded replies remain unresolved.

[PD] Build a vintage triangle by original event month and age since month-end. Estimate capture and revision patterns using independently traced source activity; a short recent plateau is not proof of finality. Keep missing records missing and test low/base/high capture assumptions. Rubin’s ignorability conditions are assumptions to assess, not a license to treat missing submissions as random.

[PD] Sample original submitted and unsubmitted line-periods across owners, value types and ages with recorded inclusion probabilities. Blind duplicate coders to workflow assignment and ultimate claim decision. Report weighted confusion tables, prevalence, raw agreement and cluster-bootstrap kappa; adjudicate against independent event and ledger evidence. An illustrative simple random sample of 400 gives worst-case binary-agreement precision near ± 4.9 percentage points at 95%, not kappa power or clustered precision. Test invariance across vintages and business contexts.

Basis: proposed measurement validation [PD]; source interpretation [HR/AI]; Rubin (1976), Government Efficiency Framework (2025), Cohen (1960).

TECHNICAL ANALYSIS

Evaluate the decision before crediting the programme

[PD] A shadow evidence benchmark can test claim accuracy under known constructed truth. A field programme trial must separately identify expansion effects on original eligible activity.

[PD] For the compact benchmark, freeze all original carryover dossiers, current amounts, permitted evidence and decision deadline. Randomize reviewer domains to current-actuals-only or RIEC, retaining support-aware current-evidence rules, full review, fixed reports and unsized responses. Lock reports at intake plus 48 hours; reveal the controlled reference at intake plus 14 days. Both potential outcomes are known only because the fixture defines them. A real adjudicator cannot observe both from one realized record.

[PD] Estimate reduction in priced reporting loss per original dossier, with added review cost on that denominator. Separately constrain false positive incremental credit among all original dossiers, including unresolved ones. This is an assignment effect on reporting decisions, not verified operating savings. Price unnecessary deferral, underclaim and overclaim; more precise claims can be economically inferior to cheaper conservative rules.

[PD] A later field study would compare an approved expansion with maintained usual service in independent resource-sharing clusters, using the same prospective population and original outcome deadline. Estimate ITT net value $E[\text{cost}(\text{continuation}) - \text{cost}(\text{expansion})]$ after all added costs and spillovers; retain a customer-service harm guardrail. Do not withdraw necessary maintenance to manufacture a comparator. Randomize where feasible; otherwise specify staggered group-time effects, comparison trends, equivalence and confounding sensitivity.

[PD] Prespecify baseline service age, reporting maturity, demand mix, capacity and price basis as modifiers. Realized submission and recognition are post-assignment variables, not controls for total ITT. Track shared reviewers and resources. Selecting low-realization lines invites regression to the mean: assume $E[\text{follow-up} | x] = m + p(x - m)$, where x is selected baseline and m population mean. Illustrative $p = 0.4/0.7/0.9$ implies 60/30/10% reversion of the selected deviation; none is historically estimated.

[PD] Calibrate domain variance, reference errors and prices before recruitment. Compare 80% MDE with break-even cost in identical units. Check actual finite null, economic-boundary and safety-boundary tails where tractable; normal planning is not exact coverage. A larger sample can fix imprecision but cannot make a dominated review rule worth buying.

Basis: proposed evaluation [PD]; illustrative sensitivity [AI]; Hernán and Robins (2016), Hudgens and Halloran (2008), Callaway and Sant'Anna (2021).

TECHNICAL ANALYSIS

Use evidence that can change the commitment

[PD] RIEC is an untested reporting and evaluation contract. Calculation validity, decision suitability, scoped approval and realized programme effects require separate dispositions.

Proposition	Prespecified rejection or hold rule
P1: capture is distinct from activity	Hold operating attribution when only posting capture is verified. Flat traced activity alone does not prove a zero causal effect.
P2: increment uses the future comparator	Reject a stated causal increment based solely on subtracting prior booked actuals; hold sizing until continuation assumptions are explicit.
P3: evidence improves the decision economically	Hold purchase unless priced reporting loss and false-credit gates clear against relevant stronger comparators.
P4: results transport across contexts	Hold extension when prespecified lag, service-age, capacity or valuation strata invalidate measurement or reverse comparative value.

[AI] Rubin motivates explicit missingness assumptions; Richardson and colleagues distinguish identification ranges from statistical precision. Target-trial reasoning fixes eligibility and the compared strategies. The Magenta Book separates monitoring from causal evaluation; efficiency guidance makes costs, periods and value types visible. Deming cautions against reacting to sparse movement. Campbell/Goodhart warns that a booked-credit target can reward capture rather than added value. Ittner/Larcker tests the measurement-to-decision chain; Tucker/Edmondson motivates repairing recurring reporting workarounds. A weighted readiness score cannot replace a missing counterfactual or scoped authority.

[HR/RC] The frozen inventory has 31 diagnostics: one weak contextual endpoint pair, 12 version comparisons, nine plan/comparator pairs and nine profiles. The weak pair has no fitted X/MR crossing, but its scope and actual/forecast basis remain unresolved. No stable phase or repeated monthly-actuals panel is supplied. Laney is neither eligible nor computed. Tentative XmR constants are 2.66 and 3.267; mechanical profiles are not operating signals.

Limitations and contribution

[HR/AI] Exact Finance-response originals remain unrecovered. A related planning workbook was read but cannot authenticate this response, sending or acceptance. Reported drafting, disclosure and escalation support a provisional registry-level hoop, not sole authorship or realized savings. No downstream reporting-cadence effect is authenticated. Capture restrictions, continuation, reference quality, prices, capacity and interference need calibration. The support-aware rule depends on an invented admissible state space; its advantage is not a field result. Unknown value is neither zero nor unmeasured upside. One-cycle results do not settle long-term consequences.

Basis: falsifiable proposed framework [PD], source boundaries [HR/RC/AI], fresh primary methods and case-specific applications. Historical numerical coordinates remain private.

Test continuation evidence after the actuals are already complete

[PD] RIEC is a shadow incrementality-reporting replay. Common current-actuals evidence is held equal, so the experiment does not mistake better record capture for programme benefit.

[PD] Each original dossier has baseline $B = 12$. Future continuation erosion D is 0 or 4 with equal probability; incremental G is -2, 0 or 4 with probabilities 0.20, 0.50 and 0.30, independently. Future amount $Y = 12 - D + G$ is verified in this constructed law. The cheap actuals-only rule reports $Y - 12 = G - D$.

Evidence law	Clear / ambiguous	Corroborated
Continuation evidence resolved	0.90 / 0.60	0.98 / 0.85
Correct given resolved	0.97 / 0.85	0.995 / 0.98
Clear-domain probability	0.70; stress 0.30	0.30

[PD] RIEC uses the reported erosion D_{hat} : report $Y - (12 - D_{\text{hat}})$ if resolved, or remain unsized. A wrong label flips 0 and 4. Randomize independent review domains, each with four original dossiers sharing evidence grade. Equal access, authority and original denominators apply; all review cost is paid. Common actuals and calibration budgets are shared, not free.

NO COUNTERFACTUAL ORACLE IN THE FIELD

[PD] Reports lock at intake + 48 hours; hidden constructed reference G is released by intake + 14 days and never tunes decisions. A real reviewer cannot observe both potential outcomes by reconciling records. Field use requires independently identified continuation evidence; known mandatory failures remain unsized.

[PD] Erosion is a future continuation state, not a missing report. $D \geq 0$ makes actuals-only conservative in this toy; continuation growth can invalidate that property. No live recognition, funding, maintenance withdrawal or customer change occurs.

PROPOSED REPORTING TEST / 2 OF 4

Detecting break-even does not mean false-credit risk is resolved

[PD] Primary ITT is reporting-loss reduction per original dossier. A separate noninferiority gate limits the added probability of reporting a positive increment when $G \leq 0$.

Domains per arm	80% MDE	Price K	BE zero power	Joint floor
Small replay / 16	0.982	0.50	29.1%	0.0%
Base replay / 128	0.350	0.50	98.0%	19.6%
Ambiguous continuation / 128	0.352	0.50	97.8%	0.0%
Corroborated replay / 384	0.202	0.90	>99.9%	>99.9%

[PD] MDE and price are judgment loss units per dossier. Economic clearance requires the reduction lower bound above K. False-credit increase must have its upper bound below an illustrative 2.5-percentage-point margin across all original dossiers; this tolerance requires decision-owner agreement.

[PD] Fixed normal radii use exact shared-grade moments: a variance ceiling for loss and an actual changed-label NI-boundary law for safety. Nominal directional error is 2.5%; joint floors are approximate. Base break-even detection is strong, yet its finite full-law joint floor is only 17.19%. A zero floor is uninformative, not zero joint power.

CORROBORATION CHANGES THE POLICY, NOT JUST THE SAMPLE

[PD] The 384-domain-per-arm candidate requires 3,072 hypothetical dossiers. Its full-law finite floor exceeds 99.9%, but recruitment and field validity remain unestablished. Selected finite loss-null errors are 2.44644% / 2.46059%; NI-boundary errors are 1.23805% / 2.18018% for base / corroborated. These checks do not establish uniform coverage or justify purchase.

[PD] Four direct allocations; no minimum-sample search. Twenty-four selected finite marginal tails use polynomial convolution, not Monte Carlo. The support-aware current-evidence rule is the stronger purchase comparator.

PROPOSED REPORTING TEST / 3 OF 4

Test the evidence already available before buying more review

[PD] $Loss = 2 \cdot \max(\text{report}-G, 0) + \max(G-\text{report}, 0)$. An unsized response costs 2.5 for this decision cycle. These are declared decision judgments, not cash or observed programme effects.

Total loss / dossier	Base	Ambiguous	Corroborated
Actuals-only / best fixed zero	2.0000	2.0000	2.0000
Support-aware current evidence	0.6000	0.6000	0.6000
RIEC including review	1.2504	1.7016	1.2577
Full review including cost	0.8794	0.8794	0.8794
Always unsized	2.5000	2.5000	2.5000

[PD] All eight fixed even reports from -6 to 8 and all 59,049 mappings from common Y to those reports or unsized are checked. Only Y=12 is ambiguous: report zero there; the other Y values identify G under the declared support. This assumption-dependent rule has no added query cost or false credit; it is not a field oracle.

[PD] Full review assumes resolution 0.98, accuracy 0.995 and cost 0.80; RIEC costs 0.50 or 0.90. Full-review cost 0.40 yields total 0.4794 and changes the preferred rule. Unsized price 1 makes always-unsized beat stressed and corroborated RIEC. Prices need independent agreement. Base net versus actuals-only = $a \cdot h \cdot 1.9082 - 0.6586 - 0.50$: a is realizability, h non-overlap; added category losses and work stay fully priced.

Net units / dossier	h = 0.50	h = 0.75	h = 1.00
a = 0.25	-0.920	-0.801	-0.682
a = 0.50	-0.682	-0.443	-0.205
a = 1.00	-0.205	0.273	0.750

PRECISION CANNOT REPAIR THE PURCHASE COMPARISON

[PD] The common-evidence rule wins in all three contexts at the primary prices. It was optimized on the assumed support and prior; budget their calibration, freeze the rule and independently test it before buying RIEC.

[PD] Reference reassignment within $G=-2/0/4$ on epsilon of original dossiers per arm permits loss-contrast error up to $40 \cdot \epsilon$ and false-credit error $2 \cdot \epsilon$, with verified costs fixed. Other support or uncertain eligibility requires new bounds.

PROPOSED REPORTING TEST / 4 OF 4

Buy information about the uncertainty that changes the decision

[PD] A separate one-time research decision concerns a reporting gap, continuation uncertainty or neither. Its value is not added to recorded recovery or programme benefit.

[PD] Prior probabilities are 0.45 / 0.35 / 0.20. Choose a reporting study, continuation study or defer. Payoffs across those states are (4,-1,0), (-1,5,-1) and (0,0,0), including downstream action costs. Each wrong signal is equally split between the other states.

Evidence scenario	Correct signal	EVSI	Cost	Net EVSI
Low	40.0%	0.10500	0.35000	-0.24500
Base	65.0%	0.93625	0.80000	0.13625
High	90.0%	1.76750	1.60000	0.16750

[PD] All 27 signal-to-action policies are evaluated per scenario. No-information value is 1.45; EVPI is 2.10. The low-information purchase loses value after cost. Low/base/high are alternative probability and price judgments, not confidence intervals or findings from the historical record.

START WITH A BOUNDED RECONCILIATION

[PD] Recover original-cohort actuals and accepted recognition definitions, register the sizing question, and identify the continuation comparison. Calibrate errors, costs and the owner-approved false-credit margin before considering a larger replay. Preserve the choice to defer research when its incremental value is below its price.

[PD] Shadow reporting decisions are not programme interventions. Even perfect recovery of actuals cannot empirically identify an unobserved continuation outcome without a separate comparison design.

REPRODUCIBILITY / READER PROTOCOL

Rebuild the event, the booking and the counterfactual

[AI] RIEC makes the inferential gap visible: an amount can be recorded correctly yet fail to measure the incremental value of the proposed next action.

Layer	Reproduction requirement
Source and status	Verify selected rows and source hashes. Preserve draft versus submitted status, unconfirmed escalation, alternative extracts and the limits of original recovery.
Comparable financial quantities	Reconstruct each numerator and commitment with its event period, forecast horizon, scope and recognized basis. Do not pool nested lines or treat a forecast as a realized result.
Recognition bridge	Link operating events to submission and booking vintages. Separate late records, reversals, scope changes and genuinely new activity before interpreting a reported increase.
Incremental comparison	Define the future population under continuation, maintenance, withdrawal and expansion. Reproduce the precise contrast and incremental cost appropriate to the decision.
Decision and information value	Rebuild the hypothetical state, evidence, action and loss law. Preserve unresolved dossiers, review costs and reference uncertainty. Compare a support-aware rule using common evidence, fixed reports and fuller review.

[PD] An independent analyst should be able to reproduce both the financial reconciliation and the point where the inference stops. A reporting-recovery scenario is not a correction to unknown historical delivery. An expansion claim must survive a credible continuation scenario; maintenance value requires its own alternative.

[AI] Raw amounts, people, source titles and version details remain in the private archive. Public quantities are independent hypothetical constructions. The archive includes the exact selected evidence, executable arithmetic, assumptions, primary-method applications and review dispositions; it does not supply missing historical approvals or create an unobserved counterfactual.

Protected evidence S1; exact workbook transcription checks; independent arithmetic and case-specific evidence, methods and decision reviews.

REFERENCES / 1

Primary sources and their role in this case

[AI] These sources support defined methods and distinctions. Historical source authentication and the incremental effect of the programme remain separate evidence requirements.

NIST/SEMATECH. Individuals Control Charts, e-Handbook section 6.3.2.2. Assess compatible event-period observations before charting a process. A YTD recognized amount and a full-year forecast are different estimands, not adjacent points in one series.

Laney, D. B. (2002). Improved Control Charts for Attributes. *Quality Engineering*, 14(4), 531–537. Require observed binary counts and stable denominators before using a rate chart. A financial commitment is neither a count denominator nor an independent sample size.

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20, 37–46. Independently classify event period, recognition vintage, missingness reason and cost basis. Agreement on coding does not establish completeness or a valid counterfactual.

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. A cohort selected for low recorded realization may improve through reversion or better reporting. Freeze selection and retain an untreated comparison before attributing movement to the operating programme.

Callaway, B., & Sant’Anna, P. H. C. (2021). Difference-in-Differences with Multiple Time Periods. *Journal of Econometrics*, 225, 200–230. For later staggered adoption, define each cohort and compare group-time effects against an eligible not-yet-treated group. Check pre-trends, anticipation and differential reporting changes.

Collier, D. (2011). Understanding Process Tracing. *PS: Political Science & Politics*, 44, 823–830. Grade analytical disclosure, draft controls and escalation separately from Finance measurement, accepted sizing, programme ownership and realized benefits.

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 information by whether it changes the next decision under the same action set and cost model. Keep research value separate from incremental programme benefit.

Primary author, institutional, journal and official-method sources. The private research record states access scope, application and limits.

REFERENCES / 2

Primary sources and their role in this case

[AI] These sources support defined methods and distinctions. Historical source authentication and the incremental effect of the programme remain separate evidence requirements.

The W. Edwards Deming Institute. The Funnel Experiment. Changing controls in response to noisy reporting can create unnecessary operational variation. Establish the measurement process before treating a financial series as a performance signal.

Campbell, D. T. (1976). Assessing the Impact of Planned Social Change. A target to improve recognized benefit can reward delayed booking or denominator changes. Preserve event cohorts and record the lag between activity and recognition.

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 booked-value target can lose its connection to incremental value. Test the mechanism connecting the financial indicator to the intended operating outcome.

Ittner, C. D., & Larcker, D. F. (2003). Coming Up Short on Nonfinancial Performance Measurement. *Harvard Business Review*, 81(11), 88–95, 139. Validate the causal chain from operational change to mature outcome and economic value. Recognition alone is a measurement event; it is not evidence of a new physical effect.

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. Repeated manual recovery of missing actuals can mask a systemic reporting obligation. Assign ownership, cadence and exception resolution while continuing to test physical delivery.

OECD / European Commission JRC (2008). Handbook on Constructing Composite Indicators: Methodology and User Guide. Keep evidence of delivery, reporting completeness, recognition and incrementality separate. A combined confidence score can conceal a decisive failure in one dimension.

Rubin, D. B. (1976). Inference and Missing Data. *Biometrika*, 63(3), 581–592. doi:10.1093/biomet/63.3.581. Do not ignore the mechanism that makes monthly actuals missing without explicit assumptions; an absent submission is not a measured zero.

Primary author, institutional, journal and official-method sources. The private research record states access scope, application and limits.

REFERENCES / 3

Primary sources and their role in this case

[AI] These sources support defined methods and distinctions. Historical source authentication and the incremental effect of the programme remain separate evidence requirements.

Hernán, M. A., and Robins, J. M. (2016). Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. *American Journal of Epidemiology*, 183(8), 758–764. doi:10.1093/aje/kwv254. Define continuation and expansion strategies, original eligibility, time zero and outcomes before comparing periods or selecting reported lines.

Richardson, A., Hudgens, M. G., Gilbert, P. B., and Fine, J. P. (2014). Nonparametric Bounds and Sensitivity Analysis of Treatment Effects. *Statistical Science*, 29(4), 596–618. doi:10.1214/14-STS499. Separate uncertainty from missing counterfactuals from sampling error; narrow a benefit range only by stating and testing additional restrictions.

HM Treasury and Evaluation Task Force (2026). *Magenta Book: Central Government Guidance on Evaluation*. Updated 15 May 2026, sections 1.7.1 and 4.2. Pair benefits monitoring with a distinct counterfactual evaluation and plan original-frame data, responsibility and collection burden before rollout.

HM Treasury and Government Finance Function (2025). *The Government Efficiency Framework*. Updated 24 November 2025, sections 4.4–4.8 and 6.2. Bridge baseline, counterfactual, outturn, delivery costs and reporting period; distinguish cash release, capacity value and a timing shift.

HM Treasury (2026). *The Green Book*, sections on business as usual, maintenance and sunk costs; updated 5 February 2026. Specify the credible future continuation path, including maintenance and opportunity costs. Compare expansion against that path rather than against a past booked amount. Applied as an appraisal analogy; no public-sector discount rate or compliance rule is transferred.

Hudgens, M. G., & Halloran, M. E. (2008). Toward Causal Inference with Interference. *Journal of the American Statistical Association*, 103, 832–842. Original manuscript. Randomize the shared reporting or operating domain to address spillovers between owners, templates and maintenance resources.

Primary author, institutional, journal and official-method sources. The private research record states access scope, application and limits.

FINAL REFEREE REPORT / INTERNAL REVIEW

Accept the disclosure; test the increment before committing it

[AI] Author-led, agent-assisted review; not external peer review. The contribution is a reconciled realization record and an explicit contract for evaluating the next incremental claim.

Major issue	Resolution and decision boundary
Financial quantities mixed	Separate recognized YTD amounts, full-year forecasts, commitments and nested portfolio context. Arithmetic ratios do not make those quantities comparable.
Missing reporting overinterpreted	A missing-actuals explanation establishes a measurement concern. It neither proves physical failure nor establishes that delivery was near plan.
Prior year used as counterfactual	Evaluate the same future population under continuation and expansion. Price maintenance versus withdrawal separately; carryover is not automatically valueless.
Draft action treated as implementation	Distinguish recorded disclosure, proposed monthly ownership and escalation text from actual submission, sustained operation and an accepted Finance ruling.
Decision model overclaim	A support-aware rule using common evidence beats paid review at the locked prices. Lower full-review cost reverses that comparison. The assumed state support needs field validation; information value is not physical programme benefit.

Partner critique: can a COO act on page one?

[AI] Yes. Require an event-to-booking reconciliation before interpreting a reporting recovery as operating improvement. Ask the accountable Finance and programme owners to specify the future continuation baseline and size only the supported incremental scope. Preserve the option to submit an explicitly unsized claim when the decisive evidence is missing.

Remaining limits and minor issues

[AI] Original close records, reporting completeness, authenticated final sizing and future counterfactual outcomes remain incomplete. The technical limitations govern the conclusions. The proposed reporting law can be replicated within its declared support; that is not proof of field performance, actual recognition recovery or causal savings.

REVIEW DISPOSITION / PASS WITHIN STATED SCOPE

[AI] Publish the bounded reconciliation and testable incrementality design. Maintain the separation between disclosed risk, Finance recognition and realized incremental operating value.

Publication requires final source, methods, numerical, partner and actual-page review records. This internal disposition is not external assurance or endorsement.