

OPERATING CASE STUDY / STRATEGY & CAPITAL ALLOCATION

Customer Loss Goals & Investment Proposals

A mandate, a funded forecast and a commitment are different claims

An operating-project case on preparing an annual quality-goal intake. It examines competing target versions, incompatible metric grains, modeled headroom and a disclosed funding gap without turning a draft into an approved or attainable commitment.

GOAL ARCHITECTURE & DECISION ASSURANCE

BLUF / BOTTOM LINE UP FRONT

Preserve the mandate. Validate the measurement and funded path separately.

ANALYTICAL FINDING

A mean funded forecast does not establish the probability of attaining a goal. Matching marginal attainment for separate goals does not establish joint attainment, and a shipment-rate improvement can coexist with a worse unit rate.

CONTRIBUTION IN THE SOURCE RECORD

Provisional registry-level hoop: compiled goal definitions and disclosures; upstream mandates, approval and delivery remained with others.

EVIDENCE STATUS

Original intake versions unrecovered. A related planning workbook was inspected; submission, adopted targets and outcomes remain unestablished.

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

STRUCTURED ABSTRACT

Make the commitment auditable without declaring it attainable

[AI] A goal package can expose the right unresolved questions while still failing to identify the adopted target, the executable funded path or the chance of reaching it.

Element	Decision-relevant conclusion
Context [HR]	A planning lead reportedly prepared a quality-goal intake under an upstream mandate, with metric definitions, funding disclosures and named open items. Central submission and reviewer acceptance are unconfirmed.
Objective [AI]	Separate the target of record from a forecast, an aspiration and the authority to change any of them. Preserve original measurement grain and decision ownership.
Approach [RC/AI]	Reconcile competing baseline/target versions; inspect entity and clock definitions; distinguish total metric headroom from incremental programme value; test what the available evidence can support.
Findings [AI]	Related workbook cells and the narrative disagree on a denominator. Mean funded reach is not a hard limit. Separate goal probabilities do not determine the chance of meeting the set together.
Decision value [PD]	The goal owner selects an authorized version; the metric owner resolves the definition fork; the funding owner supplies an uncertainty-aware delivery plan. The compiler records those rulings rather than choosing convenient values.
Implications [PD]	Use the proposed Goal-to-Outcome Reconciliation Contract (GORC). In a synthetic shadow test, a cheap staged rule beats the paid evidence rule; additional precision alone does not select the investment.

[PD] Release the mandate, funded forecast and open funding decision as separately labeled statements. A sanctioned placeholder can preserve uncertainty, but it still needs scope, accepted ownership and a closure condition. It is not proof that a value exists nowhere.

[HR] historical record; [RC] recalculated; [AI] analytical inference; [PD] proposed design. Protected sources S1/S2. Historical quantities, indexed coordinates, internal identifiers and personal information remain private.

EXHIBIT 1 / GOAL AUTHORITY

Resolve the version before asking whether the target is ambitious enough

[AI] A later file does not automatically override an earlier goal, and a worksheet prepared for a reviewer is not a confirmed central submission.

Claim	Evidence needed	Unsafe shortcut
Mandate	Issuing authority, intended scope and approved interpretation.	Treat the compiler as the target setter.
Baseline	Definition, full observation window, maturity and extract vintage.	Swap a current-year extract for an estimated close without relabeling.
Target version	Approved baseline, transformation and change decision.	Adopt the later or easier number solely because it exists.
Filed commitment	Submission receipt and accepted reviewer version.	Promote a paste-ready handoff into confirmed filing.

[HR/AI] The reviewed descriptions carry competing goal versions and unresolved baseline differences. They also describe a limited rewrite that did not cover every goal. A partial rewrite cannot be treated as the authoritative revision of a target it never changed. The artifact sequence provides a version question, not its answer.

[HR/AI] The lead assembled definitions, source references and disclosure text. Some goal wording came from the relevant owner; mandates and important measurement or financial rulings belonged to other roles. No evidence in this case authenticates a final accepted submission. Keeping this boundary does not diminish the value of a clear handoff; it prevents that value being misclassified.

ONE LOCKED DECISION RECORD

[PD] Record which authority selected the target and baseline, why the version changed and which conditions remain open. Preserve the superseded alternatives privately. Do not average incompatible versions or silently relax a mandate to match a forecast.

[HR/AI] Source descriptions are derivative, and exact intake/handoff originals remain unrecovered. A related planning workbook is a separate version, not automatic confirmation of the described intake.

EXHIBIT 2 / MEASUREMENT GRAIN

A better shipment rate can conceal a worse unit outcome

[AI] The entity named in a rate determines the result. A common label such as defects per million does not make shipment and item measures interchangeable.

Independent constructed cohort	World A	World B
Fixed original shipments / items	100 / 200	100 / 200
Failed shipments	10 single-item shipments	5 four-item shipments
Shipment failure rate	10%	5%
Item failure rate	5%	10%

[PD] The same manifest contains 50 single-item, 25 two-item and 25 four-item shipments. Moving between these constructed worlds halves shipment failure while doubling item failure. There is no denominator change: the composition of failures changes. The example is not a historical trend or evidence that one metric is intrinsically preferable.

[HR/AI] The narrative reports a verified dispatched-shipment definition for a goal, while a directly inspected related planning version retains unit-oriented or pending-definition wording. Validation status cannot be copied from one version to another. Recover the precise metric and goal version before choosing which definition governs.

[PD] Preserve item-to-shipment linkage, original cohort, exclusions and numerator/denominator clocks. Business-unit lines may remain separate for accountability; a network rate can still be valid when compatible counts and exposure weights are available. An unweighted average or a sum across incompatible grains is not that bridge.

DO NOT INFER ONE GOAL FROM ANOTHER

[PD] Publish the two measures with explicit grains and a matched-cohort bridge. A reduction in recorded shipments, an attribution correction or a different claim policy is not automatically a reduction in mature customer failures.

[PD] Entirely invented fixed-cohort proof, independently recalculated. [HR/AI] Public historical coordinates are withheld; the recovered definition fork remains unresolved.

EXHIBIT 3 / FUNDED REACH

A funded forecast below the mandate is a decision gap, not a probability

[AI] Publishing a funding gap is useful. Treating a modeled average as an absolute ceiling can make the disclosure more certain than its evidence.

Independent funded-response law	Mean reduction	Chance of reaching 30
Reduction is always 20	20	0%
Reduction is 0 or 40, equally likely	20	50%

[PD] Both constructed plans have the same mean and the same ten-unit gap to the goal. Their attainment probabilities differ. A point forecast alone provides neither the support of possible outcomes nor the probability of meeting a target. Conversely, a target beyond a supported upper bound would be infeasible under that declared model; that stronger statement needs the bound.

[HR/AI] The source describes preserving an upstream goal while showing a funded programme ambition below it. That is a valuable separation of mandate and plan. It does not establish a calibrated response curve, a verified capacity ceiling or a linear relationship between more funding and more defect reduction.

[PD] Ask for funded scope, dependency acceptance, operating capacity, effect uncertainty, seasonality and the original measurement clock. Specify whether the forecast means an expectation, median, conservative quantile or deterministic capacity bound. A target change and a funding change require their own authorities; neither should occur through an unlabeled worksheet formula.

PUBLISH BOTH STATEMENTS

[PD] Keep the goal of record and the conditional funded forecast side by side. Name the unresolved funding or operating decision. State what evidence would close the gap, not only who owns the next document.

[PD] Independent illustrative distributions, not fitted to the source. Forecast uncertainty, response heterogeneity and implementation readiness are not estimated by this case.

EXHIBIT 4 / ECONOMIC MEANING

Separate total headroom from the additional value of a funded programme

[AI] A metric opportunity and a programme benefit can occupy the same spreadsheet column while answering different economic questions.

Quantity	Required basis	What it does not establish
Metric headroom	Compatible exposure x rate change x applicable value per outcome.	That a funded intervention can achieve the change.
Programme increment	Mature effect versus a stated continuation alternative; full cost.	An addition to a parent metric opportunity already containing it.
Target share of headroom	Same cohort, endpoints, costs and period on both sides.	The probability of attainment or the best target.
Net value	Incremental benefit with overlap, adverse effects and work priced.	Cash release or Finance acceptance merely because arithmetic closes.

[HR/AI] The record explicitly distinguishes broad metric headroom from narrower programme-lane estimates and warns against summing them. Retain that warning. The descriptions also mix entity grains and maturity clocks, so a diagram of related metrics is not a proven set partition or a quantified overlap ledger.

[PD] Build an outcome-level benefit bridge before aggregation. Link each proposed effect to its cohort, counterfactual, exposure and valuation basis. Retain interactions and competing effects: an additional valid attempt can alter lateness, handling work and customer outcomes. Do not value a classification or reporting change as prevented failure without an independent outcome contrast.

A RATE IS NOT ONE CUSTOMER

[AI/PD] A rate point needs a compatible exposure denominator to become an event count. An item, shipment, concession and person are not interchangeable. Finance should validate the valuation convention and the release mechanism separately from the goal owner approving the target.

[HR/AI] Historical modeled estimates and disputed costs remain private. No implemented effect, realized saving, return on investment or payback is established.

EXHIBIT 5 / VERIFICATION SURFACE

Make another reviewer able to reproduce the same claim

[AI] A metric identifier improves traceability, but independent verification also requires accessible evidence, matching filters and a defined observation vintage.

Verification layer	Required check	Retained failure state
Definition	Named construct, entity grain, numerator, denominator and exclusions.	Unresolved version or semantic conflict.
Extract	Authorized access, matched filters, vintage and completeness.	Extract used because live verification was blocked.
Reviewer surface	Neutral link, reproducible parameters and preserved source snapshot.	Personalized or pinned view may return another population.
Closure	Owner acknowledgement and rechecked evidence after correction.	A gap list remains open even if every gap has an owner.

[HR/AI] The package descriptions report technical metric checks and a link register, but also access errors, personal parameters, filter mismatches and unresolved questions. Those limitations prevent an unqualified claim that any reviewer could reproduce every figure. An API error at one route does not prove the figure is absent from every authorized source.

[PD] Retain the exact extraction route actually used. Distinguish a source definition, a dated export and a live recalculation. When a missing baseline later becomes available, reopen the placeholder decision; do not preserve an obsolete assertion of universal absence merely because it justified the first draft.

AUTHENTICATE BEFORE CERTIFYING

[PD] Recover the original intake, handoff and current metric definition, then reproduce a bounded sample using a reviewer account. Keep the related planning workbook as its own version. Successful generation establishes output creation; separate tests must validate the transformation, forecast and approval.

[HR/AI] The source narrative and recovered related planning evidence have different provenance. Original scripts were not authenticated and replayed; access limitations are scope-specific.

EXHIBIT 6 / OUTCOME AND CONTROL

Use early signals to act without treating them as proven causes

[AI] A daily input can support a timely operating response even when the outcome matures later. That does not make the input a causal lever or the outcome ready for daily certification.

Evidence	Supported use	Extra evidence required
Lagged input association	Hypothesis and predictive check under stated timing.	An intervention contrast before claiming causal control.
Policy-adjusted line	Transparent mechanical bridge beside the recorded series.	A counterfactual policy model before claiming underlying improvement.
Early defect signal	Triage and an approved reversible response.	Original-cohort mature outcome before final effect certification.
Staffing directive	A stated operating requirement.	Published planned/actual exposure and accepted enforcement authority.

[HR/AI] The record cites lagged relationships, a composite of input measures and a subtraction of policy-related amounts from the outcome. These are not, by themselves, a randomized intervention effect or a verified no-policy customer outcome. No model in this case converts those associations into an entitlement.

[PD] Separate the action clock from the certification clock. Freeze original promises and eligibility, measure attempted actions and spillover work, and preserve all outcome categories. Determine an adequate maturity window from the actual cohort and process; a short cited completion lag does not establish universal finality across defect types or versions.

KEEP THE DECISION RIGHTS VISIBLE

[PD] The compiler can propose a cadence and document a measurement gap. Metric owners validate the measure; programme owners authorize and execute an action; Finance validates value; goal owners approve the commitment. A well-written submission cannot stand in for these decisions.

[AI/PD] Predictive association and accounting adjustments are distinguished from causal effects. No operational pilot, adopted control or mature programme result is established in the selected source.

EXHIBIT 7 / GOAL CONTRACT

Test joint assurance against a cheap staged alternative

[PD] The proposed Goal-to-Outcome Reconciliation Contract holds goal authority, measurement, funded uncertainty and joint outcome evidence in one explicit decision record.

GORC obligation	Release evidence
Lock the claim	Authorized target version, baseline and distinction between mandate and conditional forecast.
Link the measurement	Compatible entity grains, original cohort, clocks, weights and accessible source evidence.
Test funded uncertainty	Scope and response distribution, dependencies, full costs and time to maturity.
Choose the commitment	Joint attainment and priced scope/delay trade-off; maintain the conditional option and a staged comparator.
Close the governance loop	Accepted submission, recorded reviewer ruling and independently measured future outcomes.

[PD] The synthetic shadow test asks whether paired pilot observations improve a commitment choice when both goals have the same marginal attainment probability. It changes no real customer promise or funded operation. A fixed staged rule has lower priced loss and failed-assurance risk than the proposed paid rule. Better break-even detection in an expanded design still does not make economic clearance likely.

[PD] The goal pair remains the original denominator even when only one commitment is stated. The unpromised goal stays observed and forgone scope is priced. A conditional statement carries delay cost, so cautious communication is evaluated rather than automatically rewarded. An absolute risk threshold needs agreement separately from a comparison with usual practice.

THE NEXT BOUNDED ACTION

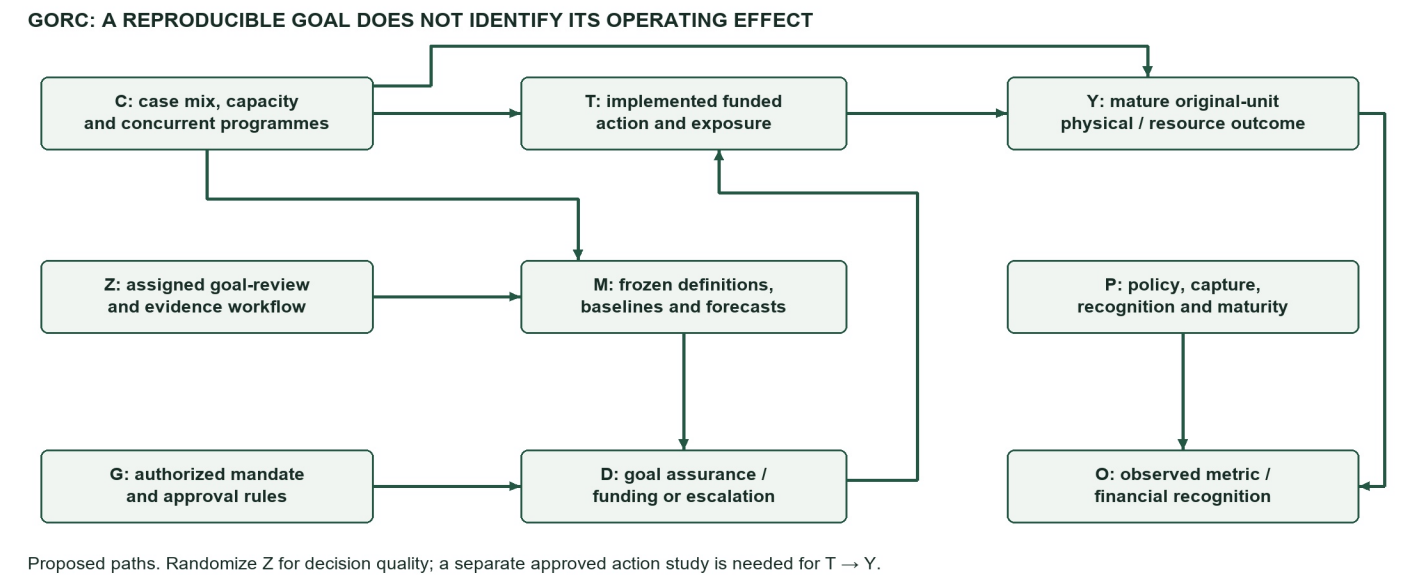
[PD] Resolve one goal-version and entity-linkage fork using the original evidence. Obtain a funded response distribution before commissioning a large shadow study. Keep the mandate intact, disclose the plan gap and ask the authorized owner to decide what changes.

[PD] All later benchmark probabilities, prices, clocks and sample sizes are invented. No achieved field power, customer effect, Finance approval or funding recommendation follows from the constructed law.

TECHNICAL ANALYSIS

Specify the promise before assessing its delivery

[PD] The Goal-to-Outcome Reconciliation Contract (GORC) separates the authorized target, its measured construct, the funded forecast and the eventual customer outcome.



[PD] Z assigns the review workflow; M holds versioned definitions, baselines and forecasts; D is assurance, escalation or a funding decision. T is actual implementation and Y its mature physical/resource outcome. Context C affects capability and risk. Policy, capture and maturity P affect observed metrics O. A better M can change a submitted goal without changing T or Y. The arrows are hypotheses, not an identified historical model.

[HR/AI] The selected evidence contains a master entry and a narrative describing the same intake. Their repetition does not authenticate independent originals. The account reports inherited goal text, a top-down mandate, local derivation, disclosed gaps and unresolved submission. A recovered related planning artifact retains a different process-metric denominator and pending identifiers. Hold the definition fork until the exact intended version and approval are authenticated.

[AI] A named system metric does not establish open access, correct filters or a successful query. A lagged association is not a causal input effect. If a total contains a component, their association partly reflects arithmetic: for $Y=X+R$, $Cov(X,Y)=Var(X)+Cov(X,R)$. Nor does subtracting a reported policy component identify a physical improvement. Keep raw and defined adjusted series, with separate evidence for each.

Basis: proposed DAG [PD]; critically interpreted registry and related artifact [HR/AI]. NIST (2024) separates implementation, effectiveness and impact measures.

TECHNICAL ANALYSIS

A nested process rate can improve while unit harm rises

[AI] Independent hypothetical fixture: the same manifest contains 100 shipments and 200 units—50 single-unit shipments, 25 two-unit shipments and 25 four-unit shipments.

Constructed outcome	Failed shipments	Failed units	Rates: shipment / unit
Ten single-unit shipments fail	10	10	10% / 5%
Five four-unit shipments fail	5	20	5% / 10%

[AI] Every unit in a failed shipment fails in this fixture. The shipment rate halves while the unit rate doubles, even with fixed exposures and a correct shipment-to-unit map. In general, k failed shipments imply between the sum of the k smallest and k largest shipment sizes in failed units under this all-units-fail assumption. Partial failures require an additional within-shipment mapping. Conceptual nesting alone cannot select a conversion or identify prevented customer failures.

Same funded mean reduction	Future support	Chance of reduction ≥30
Constant response	Always 20	0
Variable response	0 or 40, equally likely	50%

[AI] These separate invented laws have the same expected reduction but different chances of reaching the goal. A mean forecast shortfall is not a hard capability limit; if the supplied forecast is not a mean, even this probability interpretation is unavailable. A funded plan, ambition, probability forecast and physical maximum therefore need different labels.

[AI] Baseline revisions create another distinction. A hypothetical target of 80 calculated from an estimated 100 implies a 20% reduction. If the settled baseline is 60, the same target permits a one-third increase; if it is 120, it requires a one-third reduction. Freeze the rule and revision authority rather than choosing the convenient vintage. No number in these fixtures is historical.

Basis: exact independent constructions [AI], reproduced from explicit manifests and finite support. Fleming/DeMets (1996) is a surrogate-pathway analogy; Gneiting/Raftery (2007) supports forecast assessment.

TECHNICAL ANALYSIS

Reconcile denominators, clocks and value before adding

[PD] Treat each goal as metric × eligible population × time basis × version. Distinct denominators can be related through evidence without becoming interchangeable.

Layer	Required contract
Metric identity	Counted event, unique key, unit/shipment/customer grain and exclusions.
Time and exposure	Event date, original cohort, reporting vintage, mature horizon and volume.
Target authority	Mandate, baseline version, derivation rule, approved revisions and responsible role.
Funded forecast	Executable action scope, resources, uncertainty and conditional dependencies.
Economic claim	Headroom or incremental effect; shared counterfactual, price, overlap and full cost.

[AI] A conceptual goal tree supports caution about adding claims, not a quantified overlap correction. Headroom, programme effect and recognized cash differ. Reconcile unique outcome units and causal packages before valuing them; repeated goals may describe the same outcome, different outcomes or interacting effects. Separately reported business lines can still have a valid exposure-weighted aggregate when scope and dates match. Preserve strata alongside any authorized roll-up. Averaging rates without weights is a different operation.

[PD] Audit all proposed goals where feasible, then sample linked events with known inclusion probabilities across metric, business line, multiplicity, missingness and source version. Two coders blinded to workflow and outcome classify denominator, linkage, maturity and authority. Report weighted confusion tables, prevalence, raw agreement and cluster-bootstrap kappa; adjudicate using independently opened records. A simple random sample of 400 supports worst-case binary agreement precision near ±4.9 percentage points, not clustered precision or kappa power.

[PD] Validate original-cohort receipt separately from concession access and recognition. Keep late corrections and unresolved links visible; do not delete them from the original denominator. Test whether errors vary across networks, source vintages and shipment sizes. A statement that a field was not found is bounded by the searched estate, not proof of worldwide nonexistence.

Basis: proposed measurement contract [PD]; NIST (2024), Cohen (1960), ICH E9(R1) (2019). Official and clinical guidance are method analogies, not compliance claims.

TECHNICAL ANALYSIS

Test assurance value separately from goal achievement

[PD] A shadow comparison can test whether a goal review supports better commitments. It cannot establish that a delivery programme caused a customer improvement.

[PD] Freeze every original goal packet, common permitted evidence and hard authority stops. Assign independent reviewer domains to usual review or GORC. Lock assurance decisions at intake plus 72 hours. Prespecify the following 90-day forecast window before assignment; reveal reference outcomes 30 days after that window ends. These illustrative clocks require field calibration. All staged, unresolved and blocked packets remain; never infer authority from a favorable model result.

[PD] The assurance estimands are mean priced decision loss and unconditional unsupported joint assurance per original packet. Compare fixed, staged, conditional, same-evidence optimized and fuller-review policies with equal information rights and full costs. The finite benchmark can make its future law known by construction; a field reference cannot reveal unobserved programme potential outcomes. Forecast probabilities need out-of-sample calibration and proper scores distinct from business-loss pricing.

[PD] A later approved cluster trial of an action package has another estimand: $E[Y(\text{usual}) - Y(\text{offer})]$ per original eligible unit at its original promise, with mature independent receipt, ordinary-flow harm and full resources. Cluster where staff and routing spill over. Keep policy and recognition changes in the assigned analysis; separately define any hypothetical policy-standardized contrast. No conditioning on later goal attainment, adoption or claim receipt to estimate total ITT.

[PD] Prespecify baseline capacity, case mix, source age, multiplicity and programme version as moderators. Selection of weak sites invites regression to the mean: under an assumed linear conditional-mean model, expected untreated follow-up is $m + p(x - m)$, with x the selected baseline and m the population mean. No historical p is estimated. For staggered observational rollout, inspect group-time pretrends, equivalence and sensitivity; nonsignificance is insufficient.

[PD] Calibrate cluster and serial covariance, forecast transport, reference error and recruitment before sizing. Compare 80% MDE with review break-even K per original packet; inspect finite null and boundary tails where tractable. A passing power calculation cannot rescue a dominated rule or a misleading outcome.

Basis: proposed design [PD]; Hudgens/Halloran (2008), Callaway/Sant'Anna (2021), Barnett et al. (2005), ICH (2019), Gneiting/Raftery (2007). Numerical assumptions are separately declared.

TECHNICAL ANALYSIS

Make the assurance claims falsifiable

[PD] GORC tests definition, forecast, decision and outcome claims separately. A transparent target can be useful while its funding, attainment and causal effect remain uncertain.

Proposition	Rejection or hold criterion
P1: process-rate gains imply unit gains	Reject the unconditional implication by the exact manifest; require linked unit-outcome contrasts.
P2: a funded point gap proves infeasibility	Reject without a physical bound; calibrate conditional attainment distributions and constraints.
P3: GORC adds decision value	Hold purchase unless cost and safety gates clear against strong common-evidence alternatives.
P4: assurance transports across versions	Hold extension when independent errors or comparative value change beyond prespecified tolerances.

[AI] Locke/Latham makes strategy, feedback and commitment relevant moderators rather than proof of capability. Fleming/DeMets motivates testing paths outside a proxy. Gneiting/Raftery separates probabilistic calibration from target setting. ICH forces a precise population and treatment question; NIST distinguishes measured implementation from impact. These are methodological applications, not borrowed operational effect sizes.

[AI] Deming warns against reacting to sparse immature readings. Campbell/Goodhart cautions that target pressure can change tags or eligibility. Ittner/Larcker requires evidence along the input–action–outcome chain; Tucker/Edmondson directs recurring missing-data workarounds toward system repair. Composite weights do not validate the chain or make mixed-grain rates additive.

[HR/RC] The frozen variation inventory has 60 diagnostics: five weak contextual endpoint screens, 20 versions, 19 design audits and 16 profiles. No source screen crosses fitted X/MR limits; none establishes a stable phase. Annual versus partial-year, nested-window and month/week comparisons are non-diagnostic. XmR constants are 2.66 and 3.267. Laney is neither eligible nor computed. No missing weekly observations are invented.

Limitations and contribution

[HR/AI] The account supports a provisional registry-level hoop for packaging, derivation and disclosure, not exclusive authorship of inherited goal text or the mandate. Exact completed intake artifacts and central submission remain unauthenticated; a related planning artifact preserves a definition fork. No adopted target, funded capacity, realized saving or customer effect is established. Source linkage, maturity, effect transport, prices and interference remain material. An unresolved value is not zero; a disclosed forecast gap is not a demonstrated inevitable miss.

Basis: agent-assisted internal methods assessment. [HR] historical record; [RC] recalculation; [AI] analytical inference; [PD] proposed design. All public numeric fixtures are independent hypothetical constructions.

PROPOSED GOAL-PAIR TEST / 1 OF 4

Observe the pair before declaring the pair attainable

[PD] This wholly synthetic shadow test compares assurance choices. It changes no actual goal, customer promise, funding or future physical outcome.

[PD] Each original packet concerns goals A and B. Common evidence gives each 80% marginal attainment. A hidden regime is equally likely H or L; its paired outcome law is:

Constructed regime	Both met	A only	B only	Neither
H	0.784	0.016	0.016	0.184
L	0.604	0.196	0.196	0.004

[PD] Base pilots have two earlier paired observations. Their whole-pair capture probability is 0.90 or 0.55 under a shared domain grade, with clear-grade probability 0.75. Three conditionally independent packets form a domain; domains are independent. The old regime persists to the future with probability t: base 1.00, transfer stress 0.60.

[PD] The refreshed policy has four observations, capture 0.98/0.80 and t = 0.85. Conditional on regime, earlier and future outcomes are independent. Capture hides complete pairs independently; no selective missingness or regime access is granted to the acting rule.

[PD] Usual declares both from marginal goals. The audit declares both only after at least one captured both-pass and no singleton failure; otherwise it stages A. A common 5% known identity/authority blocker permits only conditional release. All original packets and both future outcomes remain.

FREEZE THE CLAIM AND THE HORIZON

[PD] Lock the shadow choice at original intake + 72 hours. Judge attainment over the next fixed 90-day window against original targets; close the independent reference 30 days later. These invented clocks and transport laws require calibration. Attaining a goal is not itself a customer-benefit reference.

[PD] Paired probabilities have identical 0.80 marginals but different joint success. Regime is per packet; evidence grade is shared within a domain. No source probability or historical implementation is used.

PROPOSED GOAL-PAIR TEST / 2 OF 4

Detecting break-even does not make economic clearance likely

[PD] Primary ITT is decision-loss reduction per original packet. A separate gate tests reduction in the unconditional chance that a stated assurance fails.

Domains / arm	80% MDE	Price K	Zero-test power at BE	Full-law joint floor
Small / 16	Not reached*	0.20	3.8%	0.0%
Base / 128	Not reached*	0.20	10.6%	0.0%
Transfer / 128	Not reached*	0.20	10.3%	0.0%
Refreshed / 512	0.354	0.40	89.4%	4.2%

[PD] Economic clearance requires a loss-reduction lower bound above K; risk superiority requires its bound above zero. Audit is used with probability lambda per original packet, usual otherwise; K is fully paid. Fixed normal radii use exact shared-grade moments and mixture-path variance ceilings. Nominal directional error is 2.5%; this is approximate planning.

[PD] The refreshed MDE is 0.354 against price 0.40, and break-even detection is 89.4%. Yet full-law economic-clearance power is only 5.1% because mean net gain is just 0.0662. Its joint floor is 4.2%. None of these designs establishes a useful purchase test.

[PD] Selected exact finite tails give base/refreshed loss-null errors 2.52139%/2.51009% and risk-null errors 2.57357%/2.45949%. Small excesses above nominal are retained. Refreshed full-law joint marginal floor is 4.23676%; it is not an exact joint probability or uniform coverage guarantee.

REDESIGN THE QUESTION BEFORE RECRUITMENT

[PD] The expanded screen would need 3,072 packets across 1,024 domains; availability is unestablished. Retain the stronger cheap staged comparator, calibrate transport and prices, and reconsider the evidence purchase. Better signal quality and a larger sample do not remove the adverse comparator.

[PD] Four direct allocations; no minimum search or Monte Carlo. Eighteen finite marginal tails use polynomial convolution. *80% detection is not reached on the declared path. A zero joint lower bound is uninformative, not zero joint power.

PROPOSED GOAL-PAIR TEST / 3 OF 4

Price the unpromised scope as well as the failed assurance

[PD] Loss is 8 if a both-goal assurance fails; staged A/B costs 1 for forgone scope plus 4 if its named goal fails; conditional costs 2.4. These are decision judgments, not dollars.

Total loss / failed-assurance risk	Base	Transfer stress	Refreshed
Usual: both	2.4456 / 29.07%	2.4456 / 29.07%	2.4456 / 29.07%
Fixed: staged A/B	1.8300 / 19.00%	1.8300 / 19.00%	1.8300 / 19.00%
Conditional only	2.4000 / 0.00%	2.4000 / 0.00%	2.4000 / 0.00%
Proposed audit	2.2676 / 23.43%	2.3801 / 24.83%	2.3794 / 22.11%
Best same-pilot rule	2.0237 / 19.24%	2.0300 / 19.00%	2.2300 / 19.00%
Fuller review	2.5300 / 19.00%	2.5300 / 19.00%	2.5300 / 19.00%

[PD] Review costs 0.20 base/stress or 0.40 refreshed on every assigned packet. Fuller review costs 0.70, observes six pairs at capture 0.995 and $t = 0.95$; its optimal rule stages. Best same-pilot rules minimize expected loss over every feasible action at every count vector, including no capture.

[PD] Perfect future-regime knowledge, not future-outcome knowledge, lowers loss only to 1.7958: information ceiling 0.0342 versus free staged, below every review fee. At staged-scope price 2, staged loss is 2.7800 versus conditional 2.4000; at conditional price 0.5, conditional-only wins. Prices and absolute risk need independent agreement.

[PD] Base net versus usual = $a \cdot h \cdot 0.74439688 - 0.36639722 - 0.20$. Here a is realization and h non-overlap of favorable loss-category changes. Added adverse categories and all paid review remain fully charged.

Net units / packet	$h = 0.50$	$h = 0.75$	$h = 1.00$
$a = 0.25$	-0.4733	-0.4268	-0.3803
$a = 0.50$	-0.3803	-0.2872	-0.1942
$a = 1.00$	-0.1942	-0.0081	0.1780

[PD] All fixed actions and exact same-evidence optima are retained under matching context. Common intake/reference research costs are shared and separately budgeted. A conditional choice has zero failed-assurance risk, not demonstrated zero customer harm.

PROPOSED GOAL-PAIR TEST / 4 OF 4

An informative study can still be a poor purchase

[PD] This separate research decision concerns measurement reconciliation, funded-response uncertainty or neither. Its value is not added to recommendation gains or programme savings.

[PD] State probabilities are 0.40, 0.40 and 0.20. Actions are reconcile measurement, study funded response or defer. State payoffs are (4,-2,-1), (-1,5,-1) and (0,0,0), including downstream action costs. Correct-signal probability varies; each error splits equally between the other states.

Evidence scenario	Correct signal	EVSI	Cost	Net EVSI
Low	45.0%	0.13000	0.40000	-0.27000
Base	70.0%	0.98000	1.00000	-0.02000
High	90.0%	1.76000	2.00000	-0.24000

[PD] All 27 signal-to-action policies are evaluated in each scenario. No-information value is 1.40; EVPI is 2.20. Every priced experiment loses net value here, even though information changes decisions. These are alternative probability and price assumptions, not confidence intervals or observed returns.

[PD] Reference uncertainty on epsilon of original packets per arm permits contrast errors up to 16*epsilon loss and 2*epsilon assurance risk within the declared loss support 0..8, with verified costs fixed. Selective capture, wrong entity linkage or changed target eligibility needs separate bounds; the toy transport parameter does not resolve them.

START WITH THE CHEAP DECISION THAT IS SUPPORTED

[PD] Recover an original goal version, settle one metric-grain fork and obtain a funded response distribution. Do not commission a large assurance study merely because the artifact has empty fields. Keep staged and conditional releases available, and retain the mandated target separately from the shadow communication choice.

[PD] No field validation, submission confirmation, operating effect or financial approval follows. A later prospective operating study needs original customer outcomes, authorized treatment versions and independent mature measurement.

REPRODUCIBILITY / READER PROTOCOL

Reproduce the commitment and its unresolved conditions

[AI] The Goal-to-Outcome Reconciliation Contract connects a proposed target to its measurement basis, funded forecast and customer consequence. Reproduction must preserve the gaps between those objects.

Reproduction layer	Required evidence and decision test
Version and authority	Freeze mandate, commitment and reviewing decision separately. Retain the recovered draft’s definition and the register’s conflicting description; neither establishes the accepted metric or submission.
Metric and denominator	Recover numerator events, eligible exposures, business-unit membership, event time and maturity time. Link shipment events to their customer-unit outcomes before comparing rates.
Forecast and attainment	Retain the funded point forecast, uncertainty distribution, dependence across goals and scenario assumptions. A shortfall against a mean does not establish impossibility.
Economics and overlap	Separate total metric headroom from incremental programme benefits. Identify affected units once, price avoidable mature consequences and subtract implementation costs.
Prospective evaluation	Lock the goal-review protocol and comparison rules before outcomes are revealed. Preserve unanswered requests, mandatory holds, review costs and changed decisions.

[PD] Replay a revised goal with every earlier baseline still available. Show whether its implication changes because the operation changes, the denominator changes, the estimated close moves or the commitment is reset. Require the accountable approver to state which change is authorized. A table that reproduces arithmetically can still encode incompatible populations.

[AI] The private archive preserves the selected register record, protected arithmetic, recovery results, executable counterexamples and internal reviews. It does not supply missing originals or demonstrate an actual filing. The public numerical examples are independent hypothetical constructions; none represents concealed historical rescaling or realized economic impact.

Protected project register and recovery audit; recalculations; proposed GORC reproduction protocol. Historical source coordinates remain private.

REFERENCES / 1

Primary sources and their role in this case

[AI] These methods distinguish a well-defined goal from an identified effect. Their use supports the proposed evaluation; it does not authenticate the historical artifacts or establish adoption.

NIST/SEMATECH. Individuals Control Charts, e-Handbook section 6.3.2.2. Assess actual ordered observations before estimating process limits. A baseline, estimated close and future target are different objects, not successive process observations.

Laney, D. B. (2002). Improved Control Charts for Attributes. *Quality Engineering*, 14(4), 531–537. Require original event counts, exposure denominators and rational subgroups. Large rate denominators alone do not justify an overdispersion correction or a process-change claim.

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20, 37–46. Audit independently coded defect states, goal status and denominator rules. Report agreement separately from correct classification and causal validity.

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 deliberately selected group of weak-performing sites can move toward its usual level without a goal-setting effect. Freeze selection and estimate reliability before attributing tail convergence.

Callaway, B., & Sant’Anna, P. H. C. (2021). Difference-in-Differences with Multiple Time Periods. *Journal of Econometrics*, 225, 200–230. For phased adoption of the goal-review protocol, define exposure cohorts, uncontaminated comparisons and mature outcomes. Assess anticipation, pre-trends and changing population composition.

Collier, D. (2011). Understanding Process Tracing. *PS: Political Science & Politics*, 44, 823–830. Attribute drafted goals, measurement definitions and exception disclosure separately from the mandate, approval, filing and downstream outcomes.

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 evidence through the commitment decision it changes under identical feasible actions. Subtract acquisition costs and keep information value separate from operating impact.

Primary author, journal, institutional and official sources. Access scope, methodological analogies and case applications are recorded privately.

REFERENCES / 2

Primary sources and their role in this case

[AI] These methods distinguish a well-defined goal from an identified effect. Their use supports the proposed evaluation; it does not authenticate the historical artifacts or establish adoption.

The W. Edwards Deming Institute. The Funnel Experiment. Repeatedly resetting targets after noisy observations can obscure variation. Establish compatible measurement windows and decision rules before revising commitments.

Campbell, D. T. (1976). Assessing the Impact of Planned Social Change. A target can redirect attention toward its measured proxy. Preserve unscored customer consequences and unresolved measurement states when evaluating goal quality.

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 targeted measure may cease to track the underlying objective. Monitor whether denominator choices or classifications change after a target becomes consequential.

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 goal specification through operating behavior to mature customer and economic outcomes. A more precise target does not identify its 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. A one-time workaround can conceal recurring measurement or handoff failures. Record unresolved sources and test whether responsible owners close the underlying defects.

OECD / European Commission JRC (2008). Handbook on Constructing Composite Indicators: Methodology and User Guide. Separate ambition, funded forecast, evidence quality and mature outcomes. Equal weights do not remove value judgments, and a composite score cannot repair incompatible denominator families.

Locke, E. A., and Latham, G. P. (2002). Building a Practically Useful Theory of Goal Setting and Task Motivation: A 35-Year Odyssey. *American Psychologist*, 57(9), 705–717. doi:10.1037/0003-066X.57.9.705. Treat commitment, useful feedback and task strategy as conditions to examine; a difficult target alone does not establish delivery capability.

Primary author, journal, institutional and official sources. Access scope, methodological analogies and case applications are recorded privately.

REFERENCES / 3

Primary sources and their role in this case

[AI] These methods distinguish a well-defined goal from an identified effect. Their use supports the proposed evaluation; it does not authenticate the historical artifacts or establish adoption.

Fleming, T. R., and DeMets, D. L. (1996). Surrogate End Points in Clinical Trials: Are We Being Misled? *Annals of Internal Medicine*, 125(7), 605–613. doi:10.7326/0003-4819-125-7-199610010-00011. A process-rate response need not represent the full customer-outcome effect when other causal paths remain.

Gneiting, T., and Raftery, A. E. (2007). Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association*, 102(477), 359–378. doi:10.1198/016214506000001437. Evaluate frozen probability forecasts against later outcomes; keep forecast calibration separate from target ambition and decision loss.

International Council for Harmonisation (2019). ICH E9(R1): Addendum on Estimands and Sensitivity Analysis in Clinical Trials to the Guideline on Statistical Principles for Clinical Trials. Final version, adopted 20 November 2019, sections A.3.1–A.3.3. Specify population, action comparison, endpoint, summary and handling of later changes before choosing the estimator.

Schroeder, K., Trinh, H., and Pillitteri, V. (2024). Measurement Guide for Information Security: Volume 1—Identifying and Selecting Measures. NIST SP 800-55v1. doi:10.6028/NIST.SP.800-55v1. Separate implementation, effectiveness and impact; document metric scope, formula, timing, source and responsible parties.

Ordóñez, L. D., Schweitzer, M. E., Galinsky, A. D., and Bazerman, M. H. (2009). Goals Gone Wild: The Systematic Side Effects of Over-Prescribing Goal-Setting. Author manuscript, hosted by Wharton, sections on narrow goals, learning and calibration. Application: inspect outcomes outside the scored goal and retain learning objectives where the goal mechanism is uncertain. This is a design caution; the case does not establish misconduct or harmful goal-setting effects.

Hudgens, M. G., & Halloran, M. E. (2008). Toward Causal Inference with Interference. *Journal of the American Statistical Association*, 103, 832-842. Original manuscript. Assign shared planning and operational domains to account for common reviewers, resources and cross-goal interference.

Primary author, journal, institutional and official sources. Access scope, methodological analogies and case applications are recorded privately.

FINAL REFEREE REPORT / INTERNAL REVIEW

Keep the ambition; make the commitment testable

[AI] Author-led, agent-assisted review; not external peer review. The case contributes a falsifiable distinction between target arithmetic, forecast attainment and mature customer benefit.

Major issue	Resolution and decision boundary
Different goal versions merged	Retain baseline, commitment and unit-versus-shipment definition conflicts. No later timestamp, reported verification or cleaner draft establishes the approved decision.
Process rate treated as customer outcome	Preserve shipment-to-unit linkage and maturity. The constructive counterexample shows why an improved process rate can coexist with worse unit outcomes.
Planning gap treated as infeasibility	Distinguish a point forecast from a capacity bound and an attainment probability. Require a distribution and dependence assumptions before asserting reachability.
Headroom presented as programme value	Separate metric opportunity, action-package effects and mature avoidable cost. Neither nested metric labels nor monetary columns justify addition.
Draft presented as delivered impact	Attribute measurement design and disclosure within the record. Drafting, intended handoff and model performance do not establish submission or customer benefit.
More information presumed worth buying	Retain the stronger staged rule and the information ceiling under the main hypothetical prices. Sample precision cannot rescue a purchase whose value is below its fee.

Partner critique: can a COO act on page one?

[AI] Yes. Keep the stated ambition visible, require a versioned commitment and funded forecast, and assign the unresolved measurement and approval questions before treating the package as executable. Ask which customer outcome must improve when the process target is achieved.

Remaining limits and minor issues

[AI] Source recovery, acceptance, forecast calibration, cross-goal dependence and mature linked outcomes constrain the conclusion. The technical section consolidates these limits. Literature analogies and exact hypothetical arithmetic do not resolve the historical identification gap.

REVIEW DISPOSITION / PASS WITHIN STATED SCOPE

[AI] Publish the goal architecture and its testable decision design. Preserve the distinction between a documented commitment, an attainable plan and an intervention that produces customer value.

Final source, numerical, methodological, partner and actual-page reviews are required. This is internal publication review, not external assurance or endorsement.