

OPERATING CASE STUDY / OPERATIONAL EXCELLENCE

Operating Commitments & Accountability

Verify the operating promise before judging its delivery

An operating-project case on repairing accountability in a distributed delivery network. It examines a commitment audit, reconciles conflicting performance interpretations, and defines an auditable path from a reported action to a mature customer outcome.

PERFORMANCE GOVERNANCE & EXECUTION

BLUF / BOTTOM LINE UP FRONT

Count a commitment only when its target, owner and deadline are verified.

ANALYTICAL FINDING

Similar percentages can answer different questions. A target with an unresolved deadline fails a complete-commitment rule; a green network average cannot rule out a failure inside a smaller operating segment.

CONTRIBUTION IN THE SOURCE RECORD

Registry-level hoop: assigned response, reported analysis and proposed completeness gate. Other owners controlled site execution.

EVIDENCE STATUS

Design and evidence-reconciliation case. Original sources remain unrecovered; adoption, customer benefit and savings are unverified.

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

STRUCTURED ABSTRACT

Make the commitment auditable before calling the plan credible

[AI] A populated performance report is useful only if the underlying promise can be found, accepted and tested.

Element	Decision-relevant conclusion
Context [HR]	A network leader was asked to explain defect movement, identify local causes and report the improvements that operating units had committed to deliver.
Objective [HR/AI]	Establish what the available record supports about current commitments; avoid filling missing promises with plausible actions or borrowed targets.
Approach [RC/AI]	Reconcile metric families, comparison bases, cohort membership, commitment fields and event clocks. Preserve unsupported and contradictory source interpretations.
Findings [AI]	Count growth and deviation from plan are different quantities. An unresolved due date fails the strict completeness rule. Aggregate process compliance does not exclude local recovery failure.
Decision value [PD]	The cadence owner should commission a bounded system-of-record audit. Hold certification until the metric, eligible frame, accepted owner, target and due date are verified.
Implications [PD]	Use the proposed Commitment-to-Outcome Contract (COC). Test whether it reduces false certification at acceptable review cost before claiming any customer benefit.

[AI] The accountable sponsor owns the decision to adopt the gate; the measurement owner verifies definitions and versions; local operating leaders accept the resources and actions. An analyst can expose a missing field without owning the downstream fix.

Claim classes: [HR] historical record; [RC] recalculated; [AI] analytical inference; [PD] proposed design. Protected register S1.

EXHIBIT 1 / DECISION CONTRACT

Separate the promise, the action and the outcome

[PD] COC creates a chain of verifiable states. Each transition requires evidence; a completed field is not proof that the next state occurred.

State	Minimum evidence	Meaning of an unresolved state
Specified	Metric, population, baseline/version, numerical target and due date	A direction of travel is not yet a testable promise.
Accepted	Named accountable role, explicit acceptance, dependencies and feasible capacity	A named person may only be a proposed recipient.
Executed	Dated action record and actual eligible exposure	A reported meeting or drafted response does not establish delivered treatment.
Matured	Original outcome deadline, follow-up age and stable ascertainment	An early reading cannot settle whether the customer outcome improved.
Evaluated	Comparable counterfactual, costs and adverse effects	A local rate change is not automatically the effect of the action.

[HR] The source register describes an assigned response and deliverables that retained missing commitments rather than inventing them. It also proposes a completeness gate. That documentary contribution is distinct from implementation of the gate or compliance by the operating units.

[PD] Preserve unresolved entries in the denominator and identify the next evidence owner. Do not classify a truthful unknown as failure to cooperate or authorize sanctions from a missing record. The first product is a credible commitment ledger, followed by a separate execution and outcome test.

Labels: [HR] historical record; [RC] recalculated; [AI] analytical inference; [PD] proposed design. Source basis S1; see private annex.

EXHIBIT 2 / METRIC RECONCILIATION

Two similar percentages do not validate one claim

[AI] Before comparing numbers, identify the outcome, unit, eligible population, time window and reference value. Numerical proximity cannot repair a mismatch.

Question	Correct comparison	What it does not establish
Did defect volume rise?	Current eligible defect count / prior eligible defect count minus one	Change in defect risk unless exposure is also reconciled.
Did performance miss the plan?	Actual rate / planned rate minus one	Week-over-week change or performance without the intervention.
Did risk change?	Current defects / current exposure versus prior defects / prior exposure	An action effect without a valid comparison and mature outcomes.
Did the action work?	Mature outcome under assigned action versus its valid counterfactual	The historical claim is not identified by any of the preceding ratios alone.

[HR/AI] The record presents a process-bucket count increase and a total-rate excess over a planned trajectory as corroborating the same headline. They use different outcomes and comparators. Both calculations can be arithmetically correct while the claimed reconciliation fails.

[PD] Require a comparison passport beside each headline: metric definition, grain, numerator, exposure, source version and comparator. Retain both original claims under their correct names. Recompute matched-period rates only after the component counts and their eligible exposure are available.

Labels: [HR] historical record; [RC] recalculated; [AI] analytical inference; [PD] proposed design. Source basis S1; see private annex.

EXHIBIT 3 / ACCOUNTABILITY BASELINE

A widening plan gap is an accountability finding, not an identified process shift

[HR/AI] The registered actual-to-plan history contains unfavorable performance. Keeping it visible strengthens the audit; relabeling it as delivered improvement would invalidate the case.

Layer	Interpretation	Required follow-up
Actual operating series	Observed outcome on its stated population and source vintage	Use a stable outcome definition, exposure and maturity before combining observations.
Planned trajectory	A managerial commitment, not a measured operating phase	Keep later targets separate from later actuals.
Actual minus plan	Accountability against a changing benchmark	Separate movement in actual performance from movement in the target.
Tentative process limits	A mechanical screen on the few available comparable observations	Do not narrate within-limit movement as a demonstrated shift or treat limits as causal evidence.

[AI] An ambitious falling target can enlarge a plan gap even with a stable underlying process. Conversely, an easier target can narrow the gap without better service. The arithmetic decomposition must preserve both movements; a performance-management score does not become a physical outcome merely because it is reviewed weekly.

[PD] The operating sponsor should retain the original promised trajectory, record any approved revision with its reason, and display the outcome on a common basis. A revision is a new commitment version. It must not overwrite the benchmark against which the earlier period was accountable.

Protected actual and target series are recalculated in the private annex, including provisional limits and source-version branches; no internal coordinates are released.

EXHIBIT 4 / CERTIFICATION

A missing due date invalidates a complete-commitment claim

[AI] The definition governs the classification. A promising numeric target with an unresolved deadline cannot count as complete under a rule requiring a target, due date and owner.

Observed record state	Defensible classification	Decision
Target present; deadline unresolved	Partially specified	Request the missing deadline and validate the accepted owner.
Target and date present; date already elapsed	Historical commitment, delivery unresolved	Assess outcome against that date; do not count it as a fresh forward promise.
No commitment found in a message search	Not located within the searched scope	Inspect the authoritative record and its version history.
Unit absent from the compared cohort	Outside the observed assessment frame	Resolve eligibility before labeling its commitment absent.
Current accepted complete record	Eligible for certification	Retain the evidence locator and verify delivery separately.

[HR/AI] The source contains both a strict completeness definition and a looser tally that includes an unresolved deadline. It also generalizes a bounded message search toward network-wide absence. The audit should correct those classifications without claiming that a direct system search would necessarily find more commitments.

[PD] Publish two rates on a frozen eligible frame: verified complete commitments and unresolved records. Measure false certification against independently adjudicated source records. A high completion rate is undesirable if teams achieve it by entering unaccepted owners, stale dates or unverifiable targets.

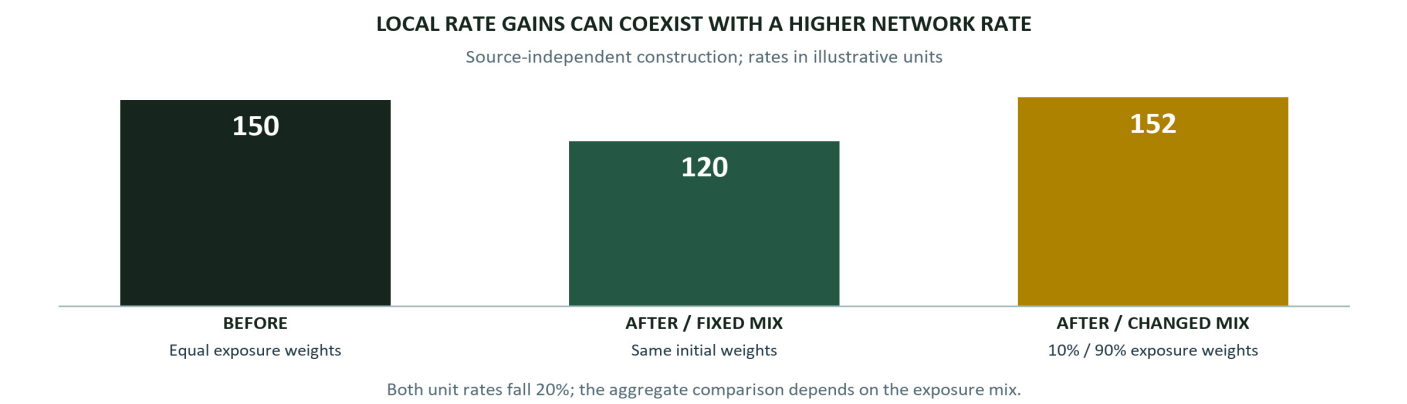
Labels: [HR] historical record; [RC] recalculated; [AI] analytical inference; [PD] proposed design. Source basis S1; see private annex.

EXHIBIT 5 / POPULATION CONTRACT

A coherent network plan requires weights and a stable cohort

[AI] Two useful local rankings need not refer to the same population. Ranking by rate, ranking by volume and excluding already-owned units create different decision sets.

Boundary	Failure if ignored	Repair
Different ranking criteria	A change of list appears to be performance movement	Declare whether the purpose is risk, volume reduction or ownership routing.
Different observation windows	Recent and trailing performance are treated as the same vintage	Align the window or retain separate comparisons.
Disjoint ownership lanes	Logical disjointness is incorrectly described as double counting	A union can be valid when grain, eligibility and time match; deduplicate only actual overlap.
Changing exposure weights	Local target achievement is mistaken for a network target guarantee	Compute the exposure-weighted network trajectory and a fixed-mix sensitivity.



[PD/RC] Independent illustration: two units reduce their rates from 100 to 80 and from 200 to 160. Equal initial exposure gives a network rate of 150. With later exposure weights of 10% and 90%, the network rate becomes 152; using fixed initial weights gives 120. Every local rate improved, yet the changing network mix reverses the aggregate direction.

[AI] This proof does not estimate the historical mix effect. It identifies the missing contract: unit targets plus exposure assumptions must reconcile to the network promise. Where future volume is uncertain, require a range of feasible mix scenarios and report the target sensitivity.

Labels: [HR] historical record; [RC] recalculated; [AI] analytical inference; [PD] proposed design. Source basis S1; see private annex.

EXHIBIT 6 / MECHANISM DIAGNOSIS

Green aggregate compliance cannot exclude a concentrated failure

[AI] Passing a process threshold describes the measured population and rule. It does not establish that all segments are healthy or that the measured process has no causal role.

Observed signal	Unresolved path	Useful next evidence
High aggregate scan compliance	A small operating cycle may concentrate residual failures	Cycle-level eligible counts, failures, recovery exposure and stable definitions.
Poor problem-solving result	Task creation, completion and physical recovery may be different events	Original eligible tasks with independent outcome verification.
A compliant upstream label	Correct scanning can coexist with later handoff failure	Time-ordered events and a validated codebook.
Null provider or cycle	A field may be populated only after an event, or may fail to populate	Schema rules, immutable events and validated join coverage.

[HR/AI] The source record includes later, related analysis showing that an aggregate view and a cycle-specific view can differ materially. That work qualifies the broad exclusion claim; it does not transfer ownership of the sibling analysis or prove the mechanism in this particular deliverable.

[PD] Audit missingness as a separate state. Compare the field-generation rule with independently verified dispatch history, and preserve unmatched units. Filling a blank by inference can manufacture an apparent diagnosis; deleting it can manufacture apparent compliance. Neither practice establishes fewer customer failures.

Labels: [HR] historical record; [RC] recalculated; [AI] analytical inference; [PD] proposed design. Source basis S1; see private annex.

EXHIBIT 7 / OUTCOME CLOCK

Test recovery disruption before excluding a later incident

[AI] A last scan before an incident does not prove that the incident had no effect on subsequent recovery. Event order narrows the hypotheses; it does not identify the entire causal path.

Clock or observation	What it establishes	What remains unresolved
Last recorded scan	Latest event in the available scan history	Whether a later incident prevented recovery or whether the history is complete.
Recorded concession date	When the financial or reporting event was recognized	The date of the initiating failure and the full follow-up distribution.
Early extract cutoff	Which outcomes were visible at extraction	Later-recognized outcomes from more recent operational cohorts.
Large share of net increase	Concentration after offsetting declines	Share of all harm, causal contribution or the most valuable intervention.

[HR/AI] The register describes a lag analysis, an incident interpretation and a proposed package-level decode. Original event linkage, the completed decode and the incident counterfactual are not established here. The correct conclusion preserves the uncertainty between initiating failure, recovery interruption and delayed recognition.

[PD] Freeze a cohort by operational event date, follow it to a prespecified mature age, and retain all original units through recovery and reporting. Test the incident against comparable unexposed units and prior trends. The operational decision is to commission the join and validate the clock before assigning cause or pricing a benefit.

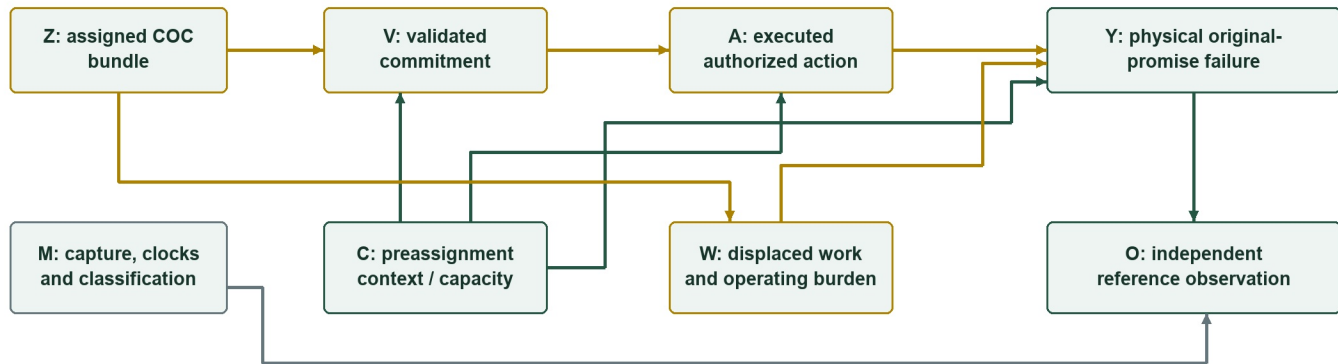
Labels: [HR] historical record; [RC] recalculated; [AI] analytical inference; [PD] proposed design. Source basis S1; see private annex.

TECHNICAL ANALYSIS

A commitment is an intermediate mechanism

[PD] The Commitment-to-Outcome Contract (COC) connects a current, accepted promise to authorized execution and an independently measured customer outcome. Those are separate claims with separate evidence.

COC: VERIFICATION, EXECUTION AND CUSTOMER OUTCOME ARE DISTINCT



C affects commitments, execution and outcomes; Z is randomized in the proposed field trial.

O depends on physical truth Y and its ascertainment M. Blinding does not make O unrelated to truth.

Do not condition the total offer effect on V, A or W; all original eligible bookings remain.

[PD] Z is the offered verification-and-follow-through bundle. V is validated commitment status, A is executed action and W is displaced work. Preassignment context C includes capacity, eligible volume, operational complexity and prior performance. Y is physical original-promise failure; O is its observed reference measure. M is capture, classification and maturation. The diagram is a proposed causal structure, not a reconstructed historical model.

[AI] A mandatory-field gate can change V and recorded closure without changing A. Extra review can also affect Y through W. The primary field comparison therefore retains all original eligible bookings and does not condition on completed forms, accepted owners, execution or subsequent classification. The shadow benchmark later in this case evaluates admissibility only.

[HR/AI] The registry reports an assigned response, delivered analysis and a proposed cadence gate. It does not establish that the gate or metric split shipped, that a package-level decode ran, or that a later customer outcome improved. An earlier adverse reporting period cannot test an intervention not evidenced as active in that period.

Basis: COC proposed DAG [PD]; bounded historical stage and timing interpretation [HR/AI]. Carroll et al. (2007); Moore et al. (2015).

TECHNICAL ANALYSIS

Local commitments need a network denominator

[AI] Exact, independent hypothetical construction. Both sites meet the same proportional rate-reduction target, but a changed exposure mix raises the aggregate. Rate units below are arbitrary; none are historical values.

Quantity	Site A	Site B	Network
Initial rate / exposure share	100 / 50%	200 / 50%	150
Later rate / exposure share	80 / 10%	160 / 90%	152
Later rate at original shares	80 / 50%	160 / 50%	120

[AI] With $R_t = \sum_i w_{it} r_{it}$, $R_1 - R_0 = \sum_i w_{i0}(r_{i1} - r_{i0}) + \sum_i (w_{i1} - w_{i0})r_{i1}$. Here the within-site term is -30 and the mix term +32. Both local rates fall 20%, yet the network rises by 2. Fixed-mix performance improves while the realized network total does not. This is a descriptive identity, not evidence of the effect of any action.

[RC/AI] The historical component-count comparison uses the prior week; the total-rate comparison uses a planned target. Numerical similarity cannot turn them into replications. Counts require matching exposure, and a component rate is not the total rate. In another independent illustration, a component falls from 40 to 20 while the remainder rises from 60 to 80: the total stays 100. Offsetting components can also make a signed contribution exceed the net change.

[PD] Freeze the outcome taxonomy, unique-event rule, eligible original cohort, forecast exposure weights and target vintage. Report local rates, observed network rates and a fixed-mix diagnostic separately. Reconcile package and line-item units before summing money or outcomes. Disjoint lanes can be pooled when these contracts match; different windows, ranks and exclusions create a different estimand, not automatic duplication.

[AI] DECISION-CHANGING RESULT

Universal local target attainment does not logically guarantee the stated network target. A commitment portfolio needs an explicit weighted roll-up and a residual-component plan.

Basis: exact independent Fraction arithmetic in private proof [AI]; source comparison-basis correction [RC]. Simpson (1951), read in original.

TECHNICAL ANALYSIS

Validate the promise and the measurement system

[PD] Audit the original site-question frame before certifying completeness. A target, dated period, accepted owner and source link are necessary evidence fields; they do not certify available capacity, achievable impact or enacted work.

Contract	Required test
Admissibility	Current metric/version; defined denominator and target period; explicit due date; independently attested acceptance.
Executability	Authorized action, necessary dependencies, resources and a documented escalation route.
Measurement	Unique outcome grain, component/total bridge, event and posting clocks, common as-of maturity and missingness.
Outcome	Independent original-promise reference, all eligible bookings retained, all-cause outcome and resource burden.

[HR/AI] Counting an unresolved due date as complete conflicts with the source’s strict rule. A mailbox search establishes visibility in that search, not completeness of every workflow system. Unchanged bridge text is not proof of unchanged physical activity. A null provider or cycle field may arise from a stamping rule; the suggested join remains a hypothesis until lineage, schema and linked entities are inspected. Earlier last scans do not exclude an incident preventing later recovery.

[PD] Draw a probability sample stratified by submitted, absent and disputed commitments, operating context and metric version. Two coders independently classify currentness, acceptance and executability while blinded to assignment and outcomes. Adjudicate disagreements against original records; report weighted confusion tables, prevalence, agreement, Cohen’s kappa and cluster-bootstrap intervals. A hypothetical simple random sample of 400 yields at most about ±4.9 percentage points for ordinary agreement at 95%; that is not a kappa-power calculation and clustering widens uncertainty.

[PD] Test invariance across regions, versions and baseline capacity. Reconcile same-ID extracts, including additions and reversals; modal lag does not close the tail. Bound unresolved dispositions. A green aggregate score cannot exclude a consequential subgroup pathway. [RC] The private inventory contains 21 source/sensitivity screens, 15 version comparisons, 6 plans and 11 profiles. No source screen has a fitted X/MR crossing; no stable phase is established. The one mechanical two-period Laney calculation has unverified binary identity and frame: zero eligible Laney panels.

Basis: source-admissibility critique [HR/AI]; measurement protocol [PD]. Cohen (1960); VanderWeele (2013), scoped primary access.

TECHNICAL ANALYSIS

Design the field test before calculating its power

[PD] After shadow validation, randomize independent cadence-and-capacity domains to COC or the usual approved cadence. Keep the operating menu and customer remedies common. This proposed field trial has no source-supported sample size or achieved power.

[PD] Freeze site eligibility and original booked cohorts before assignment; stratify on preassignment network type, capacity and baseline risk. The benefit estimand is $\Delta_Y = E[Y(\text{usual}) - Y(\text{COC})]$ per original booking, with an independent original-promise reference. Report all-cause physical failure, review/action work and ordinary-flow harm. Documentation admissibility is a separate endpoint. Capacity sharing defines clusters; map spillovers and prespecify partial-interference assumptions.

[PD] First obtain cluster-period outcome covariance, serial dependence, cluster-size variation, reference-error bounds and missingness from a blinded feasibility phase. For m independent clusters per arm with equal cluster-summary variance σ_c^2 , a rough 80%-power, two-sided 5% screen is $MDE \approx 2.80 \times \sqrt{2 \sigma_c^2 / m}$. Weighting, baseline adjustment, small-cluster critical values and unequal arms require the actual design. Calibrate low/base/high covariance and cost ranges from pilot uncertainty; test planned estimators under actual null, break-even and harm-boundary laws. No invented field simulation substitutes for these inputs.

[PD] Normalize all incremental review, operating and fixed costs to k per original booking. If c is verified avoidable cost per prevented failure and a is its realizable fraction, break-even $\Delta_{BE} = k / (c a)$. Compare MDE with Δ_{BE} and the service-harm margin before recruitment; larger samples cannot rescue negative expected value. Delay and rework are costs, not avoided customer compensation. Study feasibility and evidence value must be priced separately.

[AI/PD] Selecting high-rate sites invites regression to the mean. Under an assumed conditional-mean model, expected untreated follow-up is $m_0 + \rho(x - m_0)$, with x the selected baseline and m_0 the population mean. Hypothetical ρ values .4/.7/.9 imply 60%/30%/10% reversion of selected excess, not estimated historical fractions. Prespecify baseline moderators; never use realized execution to define the primary comparison. A necessary phased rollout needs randomized timing and calendar adjustment; an observational alternative needs justified group-time controls, pretrend plots, equivalence limits and sensitivity, not a nonsignificant-pretrend certificate.

Basis: unpowered field design and economic screen [PD]; assumed RTM sensitivity [AI]. Hemming et al. (2015); Barnett et al. (2005); Hudgens and Halloran (2008).

TECHNICAL ANALYSIS

State what would change the decision

[PD] Four propositions make COC falsifiable. A confidence interval too wide to clear a funding gate means hold for evidence; it does not establish that the effect is absent.

Proposition	Prespecified rejection or hold rule
P1: verification improves admissibility	Use all original dockets. Reject a claimed improvement if its upper confidence bound is below the promised minimum; hold if superiority or false-certification safety is unresolved.
P2: local commitments support the network target	Compute the same-vintage weighted roll-up and residual. Reject aggregate feasibility when even the approved feasible range cannot reach the target; missing weights require reconciliation.
P3: the bundle improves physical outcomes economically	Require independent mature ITT benefit above break-even with no unacceptable ordinary-flow harm. Better forms alone cannot clear this gate.
P4: implementation is feasible in relevant contexts	Audit accepted responsibility, delivered dose and capacity. Test prespecified baseline-context interactions; failure in a context narrows deployment, not the original denominator.

[AI] Harkin et al. motivate testing monitoring, not importing its effect size. Carroll and Moore separate fidelity, mechanism and context. VanderWeele explains why process proxies require outcome validation; Simpson explains why weighting changes aggregate interpretation. Deming cautions against reacting to sparse fluctuations; Campbell/Goodhart caution about targeted completion measures; Ittner–Larcker requires a tested measure-to-outcome chain; Tucker–Edmondson directs attention to workarounds and unresolved system constraints. Composite weights must express an explicit trade-off, not conceal unlike metrics.

Contribution and limitations

[HR/AI] Attribution is a provisional registry-level hoop: reported assigned analysis and escalation are supported; exclusive authorship, operating adoption and benefit are not authenticated. Delegated analytical work remains credited to its reported author. No original joined event panel, stable baseline or contemporaneous causal comparator establishes a field effect. Cohort selection, changing mix, uncertain schema, source-version conflicts and incomplete maturity limit inference. The exact roll-up proof is constructive, not a reconstruction of the historical cause. A shadow certification result does not prove target achievability or delivery improvement.

Basis: falsifiable COC propositions [PD]; research synthesis, attribution and identifying limits [AI/HR]. Full primary access scopes are preserved privately.

Verify a commitment, not just its fields

[PD] The proposed Commitment-to-Outcome Contract (COC) is tested in shadow. Certification means a current, metric-bound promise accepted by its authorized owner; it does not establish capacity, achievability, execution or customer benefit.

[PD] Assign independent review domains to form-only or COC, with 20 frozen original dockets per domain. Decisions lock at intake + 48 hours. A blinded reference by intake + 7 days reconstructs admissibility at that deadline and stays hidden from rule tuning. Original operational promises and remedies do not change. Known blockers remain unverified in both arms.

Constructed state	Share	Form complete	False certificate loss
Valid/current/accepted	40%	90%	0
Stale or wrong metric-period	25%	95%	6
Proposed, owner unaccepted	20%	85%	10
Known unresolved	15%	0%	Blocked

[PD] COC requires a complete form plus positive currentness and acceptance-evidence checks. These are fallible visible signals, not truth access. Clear/ambiguous domains occur 70/30 in base and 30/70 in stress. Conditional pass probabilities and independence assumptions are reproducible privately; no source rate calibrates them.

Price uncertainty explicitly

[PD] Retaining a valid commitment costs 3; retaining another docket costs 0.5 hypothetical resource units. Common review costs 0.10. COC adds 0.35 verification plus 0.15 setup per original docket, including blockers.

[PD] Entire benchmark is invented. Reference correctness, seven-day ascertainment and domain independence require calibration; unavailable truth stays in the denominator.

Precision must reach the investment threshold

[PD] Primary assignment effect is lower terminal decision loss per original docket. Break-even is a 0.50-unit reduction, covering all added work and setup. A separate gate requires fewer false certificates across all assigned dockets.

Scenario / domains per arm	80% MDE	Zero-test power at break-even	Joint floor at full law
Small base / 8	1.182	20.7%	>99.9%
Base / 64	0.424	91.3%	>99.9%
Ambiguous / 64	0.423	91.4%	>99.9%
Expanded ambiguous / 128	0.300	99.7%	>99.9%

[PD] These are conditional normal-planning calculations from exact finite categorical moments, with a shared domain grade and fixed law-derived variance radii. Each directional gate uses nominal 2.5%; the intersection decision needs both. The displayed joint floor is a union-bound calculation on approximate marginal powers, not an exact guarantee.

[PD] The small design cannot detect break-even reliably. The base redesign uses 2,560 original dockets. Response mixing is per docket between the two rules; it is not observed adherence. MDE uses beneficial-direction rejection of zero, while investment requires the lower loss-reduction bound to exceed 0.50.

More precision cannot repair a worse choice

[PD] Expanded sampling under ambiguity can settle the restricted COC-versus-form comparison. It cannot establish that COC beats retaining unverified commitments. Calibration and recruitment are not demonstrated.

[PD] No achieved power or sample-size minimum. Private finite-process checks report binomial intervals at selected full, zero and economic-boundary laws.

The simple safe rule can still win

[PD] Compare all policies under the same state mix, evidence rights and prices. Certification errors are decision errors; the following resource scores are neither cash savings nor customer outcomes.

Policy	Base total	Ambiguous total	Extra work
Form-only	3.441	3.441	0.00
Currentness-only	2.398	2.481	0.15
COC	1.464	1.784	0.50
Retain unverified	1.600	1.600	0.00

[PD] Currentness-only uses the same authoritative-source check; COC adds acceptance evidence. All-retain certifies nothing but pays valid-delay and review costs. COC beats form-only in both contexts, yet loses to all-retain in the ambiguous context. With valid-delay priced at 1 rather than 3, retaining also wins in base. Prices require independent decision-owner agreement.

[PD] Base favorable error reduction is 2.802; added retention loss is 0.325. $Net = a * h * favorable - added\ retention\ loss - 0.50$. Here a is realizable resource value and h the non-overlap share; every adverse category and all new work stay fully charged. Base break-even requires $a * h > 0.294$.

Net units / docket	h = 0.60	h = 0.80	h = 1.00
a = 0.50	0.016	0.296	0.576
a = 0.75	0.436	0.857	1.277
a = 1.00	0.857	1.417	1.977

Incomplete reference can reverse the verdict

[PD] If an epsilon share per arm has unknown or wrong terminal labels, worst-case contrast error is $24 * \epsilon$ loss units and $2 * \epsilon$ false-certificate units. Bounds apply to each original docket frame; they are not corrected effect estimates.

[PD] Common review 0.10 is in both policy totals. Low/base/high prices and all signed sensitivity rows are private; audit-only research cost needs its own budget.

Buy evidence only when it changes the next step

[PD] A separate one-time research decision chooses authoritative retrieval, period reconciliation or acceptance audit. Its utility is not added to the certification model and is not operational savings.

[PD] Hypothetical states are visibility gap, stale metric-period and unaccepted owner, with prior 0.40/0.35/0.25. Payoffs for the three research actions are (5,1,-1), (0,6,0), and (-1,1,8), by state. They include the downstream research decision costs; acquisition cost is separate.

Evidence scenario	Correct-state signal	EVSI	Cost	Net EVSI
Low	45.0%	0.65875	0.70000	-0.04125
Base	70.0%	2.17750	1.00000	1.17750
High	85.0%	3.08875	2.00000	1.08875

[PD] Wrong-state signal probability splits equally between the other states. Each scenario enumerates all 27 signal-to-action policies. No-information value is 2.10; perfect-information value above it is 4.00. These are alternative likelihood/cost assumptions, not confidence intervals. Negative net value means do not buy that evidence at that price.

Bound the immediate decision

[PD] First sample authoritative records, current metric bindings and owner acceptance. Measure missing truth, review time and within-domain dependence. Lock values and comparison rules before a confirmatory test. A later operating trial must separately establish execution, customer outcomes and the volume-weighted network bridge.

[PD] Independent hypothetical research-utility model. EVSI is bounded by EVPI and retains every signal and feasible action; no historical corrected decision is claimed.

REPRODUCIBILITY / READER PROTOCOL

Reconstruct the promise and its evidence, not just the chart

[AI] COC is a proposed operating contract. Its audit can be reproduced without treating private source identities or source coordinates as public material.

Layer	Reproduction requirement
Source and versions	Verify the immutable workbook, selected rows, exact claim locations, attempted original recovery and conflicting definitions. Absence of a recovered file is not evidence that it never existed.
Historical arithmetic	Recalculate component sums, actual/plan comparisons, tentative limits, alternative rates, lag accounting and implied unit costs on their original frames. Preserve unresolved grain and cost forks.
Logical analysis	Reproduce the independent exposure-mix example. Test whether the target, owner and due date truly satisfy the same completeness definition. Never fill historical blanks with hypothetical entries.
Prospective evaluation	Freeze eligibility, certification deadline and adjudication reference. Rebuild hypothetical error rates, costs, cluster variance, break-even and information value. Keep missingness and unresolved authority in scope.
Publication boundary	Historical identities and quantitative coordinates remain private. Public examples are source-independent constructions, not indexed estimates of confidential operating magnitude.

[PD] A replication succeeds when the same inputs produce the same conditional conclusion, including unfavorable findings. An apparently superior gate can fail on added review cost, stale reference data, owner disagreement or a stronger simple comparator. Reopen the decision when these conditions change.

[AI] The private annex records raw values and source paths, and the reproducibility archive carries executable analysis and final review evidence. Internal peer checks and rendered-page inspection support the publication process; they do not confer external endorsement or establish original implementation.

Protected sources S1 and source manifest; exact-calculation and independent-review files; hypothetical evaluation separately labeled.

REFERENCES / 1

Primary sources and their role in this case

[AI] These references support methods and interpretation. They do not authenticate the historical project or transfer an effect from another industry.

NIST/SEMATECH. Individuals Control Charts, e-Handbook section 6.3.2.2. Official control-chart equations. Short, selected or changing-definition histories require diagnostic restraint; a target path is not a stable operating phase.

Laney, D. B. (2002). Improved Control Charts for Attributes. *Quality Engineering*, 14(4), 531–537. Rounded rates cannot supply an authentic repeated count/exposure panel; chart suitability precedes a dispersion adjustment.

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20, 37–46. Independently code commitment validity and missingness; agreement and truth are distinct properties.

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. Selection of worst-performing units can produce untreated reversion.

Callaway, B., & Sant'Anna, P. H. C. (2021). Difference-in-Differences with Multiple Time Periods. *Journal of Econometrics*, 225, 200–230. Use group-time effects and valid not-yet/never-treated comparisons if operational adoption is staggered and observational.

Collier, D. (2011). Understanding Process Tracing. *PS: Political Science & Politics*, 44, 823–830. Assignment, documentary response, ownership acceptance and executed work require different attribution tests.

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 the decision that verification can change and deduct evidence-acquisition cost.

Clickable original publisher, author, institutional or official-method records. Source access and case relevance retained privately.

REFERENCES / 2

Primary sources and their role in this case

[AI] These references support methods and interpretation. They do not authenticate the historical project or transfer an effect from another industry.

The W. Edwards Deming Institute. The Funnel Experiment. Official exposition. Repeated escalation in response to routine noise can add instability; a control-chart screen does not identify an action effect.

Campbell, D. T. (1976). Assessing the Impact of Planned Social Change. Original institutional report. Targeting a completeness score can change reporting behavior; preserve independently adjudicated commitment truth.

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 indicator can cease to represent the underlying objective.

Ittner, C. D., & Larcker, D. F. (2003). Coming Up Short on Nonfinancial Performance Measurement. *Harvard Business Review*, 81(11), 88–95, 139. Original abstract verified through PubMed: link process measures to strategic outcomes and validate the proposed relation.

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. Distinguish local workarounds from correction of recurring system constraints; no healthcare effect is transferred.

OECD / European Commission JRC (2008). Handbook on Constructing Composite Indicators: Methodology and User Guide. Separate component definitions, weights and sensitivity; a composite score does not establish a causal priority.

Clickable original publisher, author, institutional or official-method records. Source access and case relevance retained privately.

REFERENCES / 3

Primary sources and their role in this case

[AI] These references support methods and interpretation. They do not authenticate the historical project or transfer an effect from another industry.

Simpson, E. H. (1951). The Interpretation of Interaction in Contingency Tables. *Journal of the Royal Statistical Society, Series B*, 13(2), 238–241. Application: Aggregation depends on population weights and conditioning context. The independent site-mix construction shows why local commitments need an explicit network denominator; standardization itself is not causal identification.

Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science*, 2, 40. Application: A completed form does not measure execution dose, coverage or delivery quality. Assess accepted responsibility, capacity and enacted action separately; this conceptual framework supplies no local effect size.

VanderWeele, T. J. (2013). Surrogate Measures and Consistent Surrogates. *Biometrics*, 69(3), 561–565. Application: A favorable documentation or aggregate process proxy does not certify a favorable physical outcome. Prespecify adverse direct paths and retain independent customer-outcome measurement.

Harkin, B., Webb, T. L., Chang, B. P. I., Prestwich, A., Conner, M., Kellar, I., Benn, Y., & Sheeran, P. (2016). Does monitoring goal progress promote goal attainment? A meta-analysis of the experimental evidence. *Psychological Bulletin*, 142(2), 198–229. Application: Experimental monitoring research motivates a testable monitoring mechanism. It does not establish that a mandatory commitment form caused this operating outcome or supply a transferable effect size.

Moore, G. F., Audrey, S., Barker, M., Bond, L., Bonell, C., Hardeman, W., Moore, L., O’Cathain, A., Tinati, T., Wight, D., & Baird, J. (2015). Process evaluation of complex interventions: Medical Research Council guidance. *BMJ*, 350, h1258. Application: Separate what was delivered, how it could work and the setting that enables it. Integrate process evidence with prospective outcome evaluation rather than interpreting a filled bridge as impact.

Hemming, K., Haines, T. P., Chilton, P. J., Girling, A. J., & Lilford, R. J. (2015). The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting. *BMJ*, 350, h391. Application: Use concurrent cluster comparisons when feasible. If rollout must be sequenced, randomize timing and model calendar/correlation structure; a large number of parcels cannot substitute for independent clusters.

Clickable original publisher, author, institutional or official-method records. Source access and case relevance retained privately.

FINAL REFEREE REPORT / INTERNAL REVIEW

Accept the reconciliation; test the certification rule

[AI] Author-led, agent-assisted review; not external peer review. The accepted scope is documentary accountability and prospective evaluation, not a demonstrated operating turnaround.

Major issue	Revision and disposition
Non-equivalent percentages	Separate count change, actual/plan excess and outcome risk. Numerical similarity is not independent corroboration.
Completeness classification	Apply one target/owner/date contract; keep unresolved deadlines and historical expired promises distinct.
Population and exclusion claims	Bound message-search absence; reconcile disjoint cohorts; do not rule out a mechanism from an aggregate pass rate.
Event-clock attribution	Preserve initiation, later recovery disruption and recognition delay as separate paths. No definitive incident exclusion follows from last-scan timing alone.
Contribution and benefit	Bound the leader’s role to registered assignment and analysis. Implementation, customer impact and realized savings are not established.

Partner critique: can a COO act on page one?

[AI] Yes. The next decision is a bounded authoritative-record audit with explicit ownership and an acceptance rule. The sponsor can demand verified commitments without claiming that a mandatory field fixes delivery performance. Proceed to a shadow evaluation before using certification to direct material resources.

Remaining limitations and minor issues

[AI] Original source authentication, linkage grain, reference accuracy, eligible populations, owner acceptance and post-delivery execution remain incomplete. Hypothetical cost and power inputs require field calibration. The final review accepts reproducible conditional analysis; it does not certify rollout feasibility or customer benefit.

REVIEW DISPOSITION / PASS WITHIN STATED SCOPE

[AI] The case supports a corrected commitment audit and a falsifiable design. Treat adoption and mature outcome effects as unverified until their separate evidence requirements are met.

Final disposition must be supported by source, methods, numerical, partner and actual-page reviews in the private publication record.