

OPERATING CASE STUDY / CUSTOMER EXPERIENCE & LOGISTICS

Defect Ownership & Service Recovery

Separate operational ownership, delivery evidence and customer entitlement

An anonymized operating case on an executive loss review: replacing blanket blame with specific remediation, reconciling conflicting performance claims, and defining safeguards for automated customer refunds.

PROJECT CONTEXT / LOGISTICS QUALITY, EXECUTIVE ASSURANCE & CUSTOMER REMEDIES

BLUF / BOTTOM LINE UP FRONT

Assign each fix to a verified mechanism.
Protect remedies as evidence is repaired.

ANALYTICAL FINDING

A missing delivery event does not establish a physical loss. A low refund-reversal rate does not establish correct original decisions. Reconcile the record, the event time and the authorized remedy before changing ownership or payment controls.

CONTRIBUTION IN THE SOURCE RECORD

Authored the answer and corrections. Provisional contribution; sending, acceptance and downstream execution remain unverified.

EVIDENCE STATUS

Original artifacts were not recovered locally. This is an evidence-reconciliation and operating-design case; no causal customer or financial benefit is established.

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

STRUCTURED ABSTRACT / THE EXECUTIVE DECISION

Fund the evidence and ownership gaps before changing the remedy

[AI] Operational recovery and customer remediation solve different problems. A defective custody record can create an apparent loss, but correcting that record does not by itself establish that a customer received the shipment or lost entitlement to a refund.

Element	Decision-relevant finding
Context [HR]	An executive asked why reported losses had increased and whether automated refund controls could move faster. A draft attributed most failures to station execution. A reviewer challenged the explanation and requested accountable actions.
Objective [HR]	Replace blanket attribution with a mechanism-specific answer, correct unsupported figures and route unresolved questions. The recorded author owned the answer and rewrite, not the underlying technology, refund policy or remedial programs.
Approach [HR/RC]	The work paired differentiated verdicts with an action register and claim-verification record. This case rechecks arithmetic, population compatibility, custody inference, recovery timing and the difference between decision-time error and later reversal.
Findings [RC/AI]	The source retains incompatible headline arithmetic and pooled/channel rates. Identical reversal logs can arise from very different original refund accuracy. Missing events cannot locate physical custody without a validated observation mechanism.
Decision value [PD]	Require an evidence contract for each proposed fix: unique entity, event time, population, mechanism, accountable owner and authorized customer remedy. Evaluate capture and reconciliation changes without extending the existing remedy deadline.
Implications [PD]	The operations lead owns recovery controls; the platform lead owns event reliability; the policy owner retains remedy authority. The analytical lead blocks unsupported performance claims. Finance recognizes value only after mature entitlement and incremental costs reconcile.

ACTION FROM THIS CASE

[PD] Approve an instrumented shadow evaluation of the Remedy Evidence Contract (REC). Do not approve wider refund restrictions from a lower recorded-loss rate, an owner label or a small reversal rate alone.

[HR] Source S1: selected project register; originals remain unrecovered. [RC] recalculated; [AI] analytical inference; [PD] proposed design. Identities and raw values remain private.

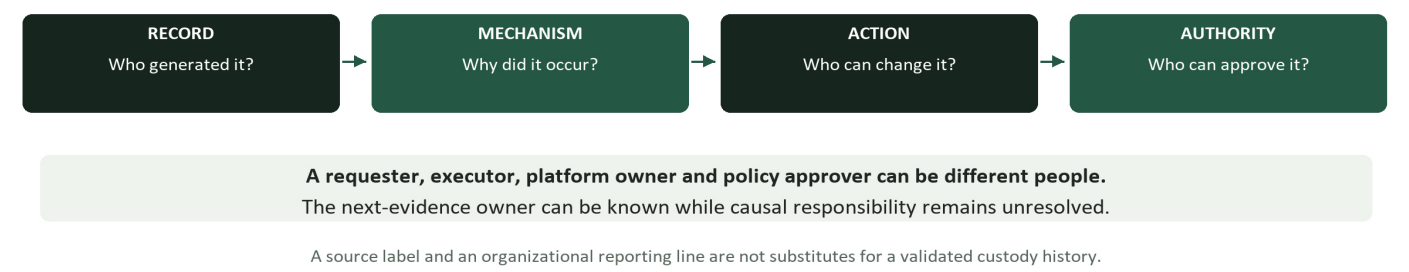
EXHIBIT 1 / DIAGNOSE BEFORE ASSIGNING RESPONSIBILITY

A failure label does not identify the team that can prevent its recurrence

[HR] The rewrite replaced one blanket explanation with several controllability verdicts. It reversed an earlier interpretation of centrally executed work, challenged a reviewer on a pre-dispatch gap and redirected a missing-event problem toward technology owners.

ONE EXCEPTION / SEVERAL DIFFERENT RESPONSIBILITIES

Proposed accountability structure / investigate the mechanism before assigning causal blame



Observed record pattern	Remediation hypothesis and evidence needed
Pre-dispatch exception	Reconcile expected load, physical handoff and exception disposition. Absence of a downstream event alone does not prove the shipment stayed upstream.
Recovery or routing gap	Trace requester, approver, executor and system permissions. A remote executor may act on a local request; neither identity alone establishes the causal origin of loss.
Custody or return exception	Use time-ordered custody evidence and physical receipt. A later return by the same worker cannot establish possession at every earlier point.
Delivery-event discrepancy	Test event capture, persistence, identity matching and actual customer receipt. A record repair can correct apparent loss without changing physical delivery.

[AI] A practical owner can be accountable for the next investigation even when causal responsibility is unresolved. Record that distinction. The proposed lead owns evidence collection and an operating response; the label is not a finding of fault or an authorization to penalize a partner.

RESOURCE THE SPECIFIC CONTROL

[PD] Name the action owner, evidence owner and policy approver separately. Require a mechanism test and a reopening trigger for each disputed attribution.

[HR] Source S1: selected project register; originals remain unrecovered. [RC] recalculated; [AI] analytical inference; [PD] proposed design. Identities and raw values remain private. Diagram is a proposed accountability structure, not proof of historical causal ownership.

EXHIBIT 2 / TEST WHETHER COMPARISONS ARE POSSIBLE

Some printed rates cannot belong to the pooled population claimed beside them

[RC] In the source, two aggregate rates lie below both of their displayed channel rates. A third matches the lower channel despite a higher rate in the other channel. These comparisons require a scope and exposure reconciliation before they support accountability.

A POOLED RATE MUST LIE INSIDE ITS COMPATIBLE CHANNEL RANGE

Wholly hypothetical rates / exhaustive disjoint channels, same outcome and period

110 / REPORTED POOL



Pooled rate = $w \times 120 + (1 - w) \times 300$, with $0 \leq w \leq 1$.

An out-of-range result fails the joint scope contract; it does not identify which input is wrong.

Rounded endpoints require an interval test. Exact equality at the minimum excludes positive higher-rate exposure.

[RC] For exhaustive, disjoint channels with the same outcome and time basis, a pooled rate is a nonnegative exposure-weighted mean. It must lie between the channel rates. In the invented example, rates of 120 and 300 permit a pooled value from 120 to 300; a reported pool of 110 is impossible under that contract.

Diagnostic result	Decision consequence
Outside the feasible interval	At least one common-scope assumption fails: population, unit, denominator, period, definition or extraction. A source error is also possible. Do not choose an explanation without the missing evidence.
At the lower endpoint	Under exact values, any positive exposure in a strictly higher-rate channel raises the pool. With rounded values, calculate the allowed high-rate weight before calling the endpoint impossible.
Different source lineages	Preserve each value with its own definition. A different lineage can explain a cross-document difference; it cannot excuse arithmetic that fails inside one claimed comparison.

THE NEW FINDING

[AI] The channel contrast cannot support a precise remediation split until its exposure basis is repaired. Recalculate within each compatible version, preserve unresolved forks and withhold the pooled causal narrative.

[HR] Source S1: selected project register; originals remain unrecovered. [RC] recalculated; [AI] analytical inference; [PD] proposed design. Identities and raw values remain private. Numerical illustration is wholly hypothetical. Raw source calculations and rounding tests remain private.

EXHIBIT 3 / DISTINGUISH DECISION-TIME ERROR FROM LATER REVERSAL

The same reversal rate can conceal very different original refund accuracy

[AI] A reversal is an observed later event. Its frequency depends on original decision quality, later delivery, evidence capture, matching, eligibility and policy. A small reversal rate cannot identify which of those mechanisms produced it.

IDENTICAL REVERSAL LOGS / DIFFERENT TRUTH AT THE ORIGINAL DECISION

Wholly hypothetical: 1,000 refunds; 20 recorded reversals; simplified receipt-based rule

WORLD A 20 already received All 20 detected Original inconsistency: 2%	WORLD B 200 already received Only 20 detected Original inconsistency: 20%	WORLD C None already received 20 received later Original inconsistency: 0%
--	---	--

A 2% observed reversal rate cannot distinguish missed errors from legitimate later state changes.

Actual entitlement requires the approved policy and independent evidence; receipt alone may not determine the remedy.

Hypothetical world	Same log; different original state
A / Complete capture	Of 1,000 refunded shipments, 20 were already received before the refund; every such case is detected and reversed. Decision-time receipt inconsistency is 2%.
B / Incomplete capture	Of the same number refunded, 200 were already received; only one in ten is detected and reversed. The log still shows 20 reversals, but decision-time receipt inconsistency is 20%.
C / Later delivery	None had been received when the refund decision was made. Twenty arrive later and reverse under the constructed rule. The log again shows 20 reversals, with no decision-time receipt inconsistency.

[PD] The construction assumes a simplified receipt-based refund rule solely to demonstrate nonidentification. In a real policy, physical receipt may not remove entitlement. Audit contractual remedy conditions independently; do not turn a receipt label into permission to charge a customer.

CHANGE THE VALIDATION QUESTION

[PD] Establish what was true and observable when the decision was due, what changed later, and which remedy was authorized at each time. Measure false payment, wrong withholding and delayed remedy separately.

[HR] Source S1: selected project register; originals remain unrecovered. [RC] recalculated; [AI] analytical inference; [PD] proposed design. Identities and raw values remain private. All counts, policies and probabilities in this exhibit are invented. No source accuracy or reversal statistic calibrates the construction.

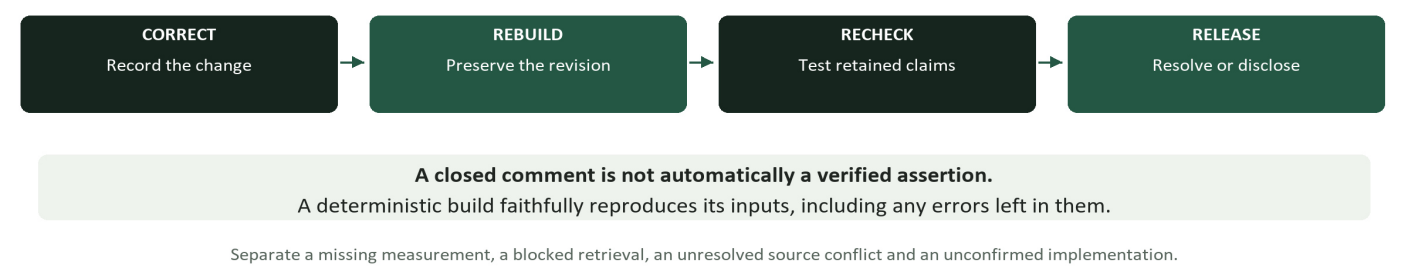
EXHIBIT 4 / SEPARATE CORRECTION FROM CERTIFICATION

A documented correction process can still leave the final answer unverified

[HR] The recorded author corrected or removed several draft figures, withdrew an earlier ownership claim and left measurement questions open. The same record shows that other source conflicts remained in the final narrative.

DOCUMENT PRODUCTION DOES NOT COMPLETE THE EVIDENCE CHAIN

Source record: corrections were made while unresolved claims and data requests remained



Evidence issue	Required disposition
Headline comparison	Recompute the printed change from its stated endpoints. Preserve the original statement and the recalculation privately; obtain the intended comparator before selecting a public headline.
Nonclosing contribution bridge	Signed contributions may exceed the net change when other components offset them. Demand the complete bridge, compatible denominators and explicit residual; do not force shares to sum by deleting offsets.
Claimed completion	A prior report that a gap was uncovered and a later claim it is closed need a dated deployment and verification record. Neither statement alone resolves the contradiction.
Open information request	Distinguish a metric that does not exist from an available trend that could not be accessed. Assign measurement creation to the former and authorized retrieval to the latter.

[AI] Reproducible document generation can reduce accidental reintroduction of recorded corrections when its rules are enforced. It does not validate the numbers embedded in the script. Release review must recheck final claims, including retained exceptions to the verification rule.

RELEASE STANDARD

[PD] Close a comment only against its stated evidence requirement. Keep unresolved claims visible in the release record; writing an answer, resolving a comment and verifying the answer are different events.

[HR] Source S1: selected project register; originals remain unrecovered. [RC] recalculated; [AI] analytical inference; [PD] proposed design. Identities and raw values remain private.

EXHIBIT 5 / PRESERVE THE CUSTOMER OUTCOME

Reducing recorded losses or refunds does not establish better delivery

[HR] The source describes customer-contact reduction after faster automated remediation and concerns about payments triggered by missing delivery records. Both are second-hand statements; neither identifies the incremental effect of a proposed tightening.

FOUR OUTCOME LAYERS / NO SINGLE METRIC CERTIFIES THE OTHER THREE

Proposed measurement architecture / fixed original population and explicit maturity



Outcome layer	What must be evaluated
Physical service	Receipt by the promised horizon, actual loss, recovery time and later delivery. An event-publication fix need not change any of these.
Record integrity	Correct unique-entity linkage, event time, posting time and evidence available at the remedy deadline. More captured events can improve visibility or introduce false matches.
Customer remedy	Correct entitlement, timely payment, wrongful withholding, erroneous reversal, customer effort and appeal resolution. A smaller contact count is not a complete customer outcome.
Economic consequence	Avoidable erroneous payments and handling cost, less added verification work, remediation, delay and implementation cost. Owed payments are not waste to be removed.

[AI] A lower payment total can coexist with unchanged physical service and worse customer treatment. Conversely, faster correct remediation can increase near-term outflow while improving the customer outcome. The decision requires joint evidence, not a single favorable financial direction.

PROTECT THE EXISTING COMMITMENT

[PD] Keep the approved remedy rule, decision deadline and access rights unchanged during shadow evaluation. Escalate uncertain evidence through the authorized path; do not introduce an extra waiting period to make the metric look better.

[HR] Source S1: selected project register; originals remain unrecovered. [RC] recalculated; [AI] analytical inference; [PD] proposed design. Identities and raw values remain private. The outcome structure is proposed; no customer effect or realized financial benefit is attributed to this project.

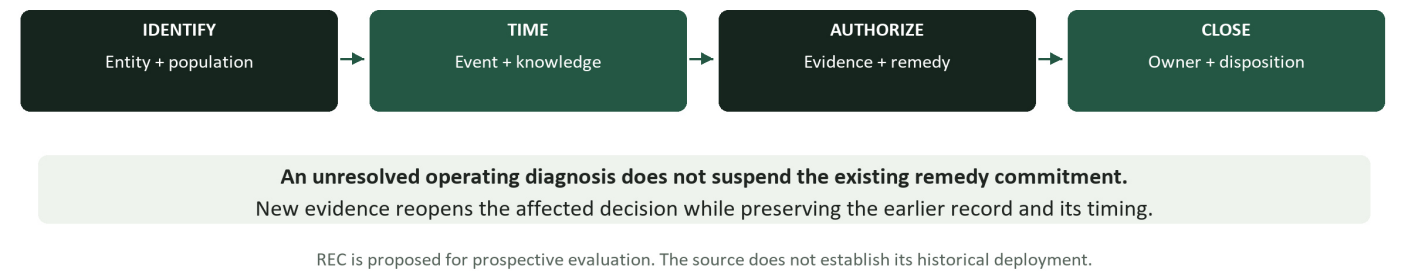
EXHIBIT 6 / INSTALL THE EVIDENCE CONTRACT

A complete remedy decision needs an entity, a clock, an authority and an owner

[PD] The Remedy Evidence Contract links an operating exception to a specific decision. It makes the evidence required to act explicit and gives unresolved cases a responsible next step without silently changing the customer promise.

REMEDY EVIDENCE CONTRACT / PROPOSED

Carry the original identity, clock and authority through the entire investigation



Contract field	Minimum evidence and accountable role
Entity and population	Metric owner freezes the original shipment population, unique identifiers, channel, eligibility and denominators. Replacements, returns and duplicate events remain linkable.
Event and knowledge time	Platform owner retains physical-event time, posting time and evidence actually available before the decision deadline. Later knowledge cannot be inserted into the earlier record.
Mechanism and remedy	Operations lead records the testable failure mechanism. Policy owner confirms entitlement, approved evidence and authorized action; ambiguous records follow the existing remedy path.
Ownership and closure	Assign investigation, execution and approval separately. Close against a verified disposition and customer outcome, preserving disagreement, appeals and reasons for reopening.

[PD] Use a frozen shadow cohort first. Retain blocked, unresolved and adverse cases. Validate the reference with blinded independent review before connecting the proposed capture and reconciliation process to customer-facing decisions.

DECISION TO SCALE

[PD] Scale only after the proposed benefit test and customer safeguards pass on the same original population and maturity horizon. Insufficient evidence leaves the expansion decision open; it does not remove an existing remedy obligation.

[HR] Source S1: selected project register; originals remain unrecovered. [RC] recalculated; [AI] analytical inference; [PD] proposed design. Identities and raw values remain private. REC is a proposed framework, not a historically deployed program.

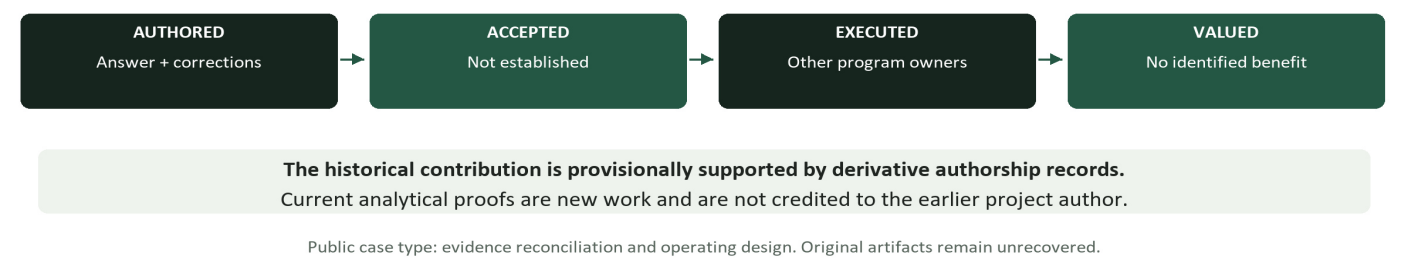
EXHIBIT 7 / ATTRIBUTE THE WORK PRECISELY

The demonstrated contribution is the analytical rewrite, not the programs it cites

[HR] The record attributes the answer, selected corrections and document-generation work to the operating leader. It assigns the original thread, source bridge, technology fixes, refund logic and downstream implementation to other owners.

AUTHORSHIP / ACCEPTANCE / EXECUTION / BENEFIT

Contribution evidence is specific to a work product and a stage



Contribution claim	Process-tracing assessment
Authored answers and corrections	Provisional registry-level hoop: the register describes named authorship metadata, a timed comment and a build script. Originals are not recovered, so authenticity and the complete revision chain remain unverified.
Revised controllability interpretation	Provisional hoop support: documented reversal and reviewer feedback fit the claimed analytical role. They do not establish that the revised attribution is causally correct.
Changed operating ownership	Unestablished execution: the answer proposes owners and an assignment for an unowned gap. Acceptance of that assignment and actual implementation are not recorded.
Delivered financial or customer impact	Unestablished. Imported avoidance estimates belong to adjacent programs; no mature counterfactual, booked benefit or approved customer-outcome evaluation is present.

[AI] The value of the case lies in making a difficult ownership and evidence problem explicit, including where the answer itself remains incomplete. The present analysis adds feasibility and identification tests; those new tests must not be retroactively attributed to the historical author.

PORTFOLIO POSITIONING

[PD] Present this as an executive evidence-reconciliation and operating-design case. Distinguish a prepared reply from a sent communication, an accepted decision, an implemented control and a realized result.

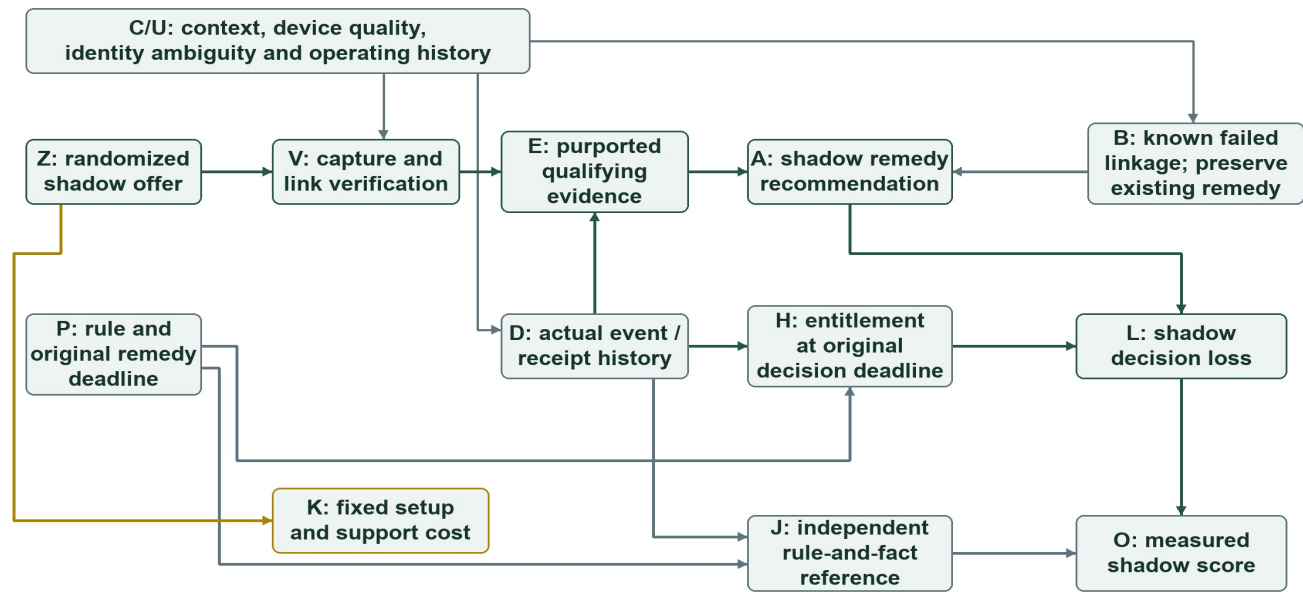
[HR] Source S1: selected project register; originals remain unrecovered. [RC] recalculated; [AI] analytical inference; [PD] proposed design. Identities and raw values remain private. Process-tracing terminology follows Collier; tests do not replace source authentication.

TECHNICAL ANALYSIS

A better evidence record need not change a customer remedy

[AI/PD] The proposed Remedy Evidence Contract (REC) separates physical events, contractual entitlement, recorded evidence and the action a policy would recommend. The source supports an executive explanation and proposed fix routing; it does not establish that a changed refund rule improves customer outcomes.

PROPOSED SHADOW STUDY: RECORD, ENTITLEMENT AND REMEDY ARE DISTINCT



K is charged once outside L. Reference J depends on actual history and the approved rule.
Shadow recommendations do not change payments, physical delivery or customer deadlines.
Known mandatory linkage failures retain the existing remedy in both workflows.

[PD] Randomized shadow workflow Z improves capture and identity verification V, affecting purported qualifying evidence E and shadow recommendation A. Actual receipt/event history D also affects E. D and the approved contract P determine entitlement H at the original decision deadline. A and H determine decision loss L; fixed implementation cost K is charged separately once.

[PD] Preassignment context C/U affects operational history and evidence quality. A known linkage blocker B requires the existing remedy recommendation. Independent reference J reconstructs D and P; its scoring role is separate from the workflow being tested. L and J produce observed score O. Reference blinding does not make ascertainment unrelated to truth.

[AI] The shadow test has no path from its recommendation to actual payments, physical delivery or customer welfare. Live remedies continue. A later authorized policy change would add those paths and require a separate identification, customer-protection and power plan. Lower shadow payment frequency alone is not a benefit.

Basis: proposed shadow evaluation [PD]; source-derived separation of event capture, refund logic and remedy authority [HR/AI]. The graph is not an identified historical causal model.

TECHNICAL ANALYSIS

The same reversal rate can hide very different decision-time facts

[AI] Consider three independent hypothetical worlds, each with 1,000 flagged shipments refunded and 20 recorded reversals. Assume each recorded receipt is genuine. The observed refund and reversal totals are identical.

Hypothetical world	Received before refund	Received later	Observed reversals
A: complete capture	20	0	20
B: incomplete capture	200	0	20
C: later state change	0	20	20

[AI] In A, all earlier receipts are captured. In B, only one tenth are captured. In C, receipt occurs after the original decision. The same 2% reversal rate therefore accompanies 2%, 20% or zero already-received shipments at the decision. These are physical-state distinctions, not automatic judgments about contractual refund entitlement.

[AI] Let r be the reversal fraction, p the pre-decision receipt fraction and s the probability that a pre-decision receipt produces a recorded reversal. The identity $r = s \times p$ requires a common original population, no false-positive reversals and no reversals caused by later receipt. Under those restrictions and known s , $p = r/s$. Unknown later receipt destroys even the proposed positive lower bound: world C has r greater than zero while p is zero.

[AI] A reported accuracy percentage cannot resolve this ambiguity without its reference definition, sample selection and full error table. Nor does one minus the reversal fraction estimate accuracy by definition. If two checks share device loss or identity-link failure, their agreement may be common failure rather than independent corroboration.

[PD] Audit original decision-time entitlement separately from eventual delivery and reversal execution. Preserve event time, evidence-availability time and policy version. Sample unreversed as well as reversed cases. A source-label change or a later delivery must not be allowed to rewrite what was knowable or owed at the earlier decision.

[AI] Hui and Walter show that estimating error without a perfect reference needs identifying assumptions. This case supplies neither independent test errors nor validated invariant error rates. The constructive worlds establish non-identification; none estimates a historical error rate.

Basis: exact independent synthetic enumeration in private/counterexample.py [AI]; Hui and Walter (1980). All numeric examples on this page are hypothetical, not transformed source values.

TECHNICAL ANALYSIS

Receipt, evidence availability and entitlement require separate clocks

[HR/AI] The register describes delivery-event publication failures, changing shipment identifiers and a reversal mechanism. Each can alter an observed financial record without identifying when physical receipt occurred or whether the earlier remedy was appropriate.

Clock or construct	Required record and interpretation
Physical occurrence	Original shipment identity, actual receipt/return event time and independent evidence of what occurred. A later return does not establish earlier custody.
Evidence availability	When each record became usable, its identity link, revision history and provenance. A posted timestamp is not necessarily occurrence time.
Decision and entitlement	Original remedy deadline, approved policy version and facts that rule requires. Receipt alone may not remove entitlement.
Payment and reversal	Actual issue, reversal, retry and customer notification times; unique financial transactions distinguished from physical units.

[AI] Refund and delivery are not necessarily competing terminal outcomes: one shipment can experience both, in either order. A state history should distinguish refunded-not-yet-received, received-after-refund and received-before-refund-with-late-evidence. Treating refund as censoring receipt would erase the very pathway under investigation.

[PD] Freeze the original cohort before capture or reconciliation changes eligibility. Join the physical entity, original and replacement identifiers, customer order, financial transaction and versioned contract with explicit one-to-many rules. Retain ambiguous links; do not multiply financial amounts across event joins or collapse distinct payment transactions merely because they concern one package.

[AI] Aalen and Johansen supply a multi-state estimation framework under specified observation and process assumptions. This source has no linked event-history panel on which to fit it. A fixed-horizon table is a useful first deliverable; transition probabilities additionally require validated risk sets and censoring assumptions. No package is presumed permanently lost simply because it lacks a later scan.

[PD] Separate decision-time correctness from eventual outcome and customer remedy delay. Later independent adjudication can discover earlier facts, but later physical events cannot retroactively change the reference question. Unknown final status remains unknown and is bounded on the original denominator.

Basis: recorded event-path and identity concerns [HR]; proposed temporal measurement contract [PD]; Aalen and Johansen (1978) for conditional multi-state methods, not an applied historical estimate.

TECHNICAL ANALYSIS

A classification rule routes investigation; it does not prove control

[HR/AI] The reply replaced a blanket operational attribution with differentiated bucket verdicts. That is a more useful hypothesis structure. The register still does not authenticate the causal responsibility of every proposed owner or close its metric-lineage disputes.

[AI] A common-scope total rate is a denominator-weighted average of its disjoint, exhaustive subgroup rates. It must lie between them when weights are nonnegative. The source contains aggregate/channel pairs that fail this condition under the claimed alignment. Exact equality with the lower channel requires zero higher-channel exposure; rounded displays may admit only a small share. Verify entity, numerator, period and exposure before using the comparison for accountability.

[AI] Headline growth also fails to reproduce from some printed baselines within a single lineage. A different metric definition can explain a different number, but does not automatically explain incorrect arithmetic inside one stated contract. Bucket counts and rates imply a different exposure from the headline. Preserve both versions and the failed checks; selecting and labeling one alone does not validate it.

[AI] Conversely, positive component contributions can exceed 100% when an offsetting component declines. That is a signed bridge, not an automatic error. It requires common scope and an explicit residual; the residual does not identify a new physical cause. Never convert an unreconciled accounting remainder into an owner entitlement.

[PD] Use three separate assignments: custodian of the recorded evidence, owner of the next investigation or feasible remedy, and any causal-prevention claim. The first two can be agreed prospectively without proving the third. A station requesting an action and another team executing it may each hold a relevant control; requester identity is not a universal causal ownership rule.

[AI] Absence of an on-road scan is compatible with a missed publication event as well as a pre-dispatch failure. A subsequent same-driver return is compatible with later recovery or handoff. Neither observation alone proves the earlier physical state. A named owner is a governance action, not evidence that its future intervention will prevent the defect.

[PD] Preserve the source's disputed weather mechanism as a hypothesis. Small coded weather shares do not rule out every indirect effect, while regional elevation and an aging narrative do not establish one. A single weekly decline and a slower-rising bucket likewise provide no controlled treatment effect.

Basis: source conflicts and reproducible arithmetic [HR/RC]; conditional weighted-mean and signed-accounting identities [AI]. Proposed ownership is distinct from identified causal prevention.

TECHNICAL ANALYSIS

Validate the remedy decision against an independent reference

[PD] The validation unit is an originally eligible ambiguous-receipt episode at its unchanged remedy deadline. The reference is contractual entitlement at that deadline, independently reconstructed from facts and the approved rule; it is not the workflow’s own delivery flag.

Measure	Reference and denominator
Evidence sensitivity	Purported qualifying evidence among independently adjudicated non-entitled episodes; preserve the original deadline.
Wrong-withholding risk	Shadow withholding recommendations among all original episodes that were entitled; report both conditional and all-episode rates.
Unnecessary-payment risk	Shadow payment recommendations for non-entitled episodes, separately from actual payments.
Coverage and unknown truth	Known linked and independently resolved shares on the original cohort; unresolved cases are not silently removed.

[PD] Independently double-code a random core of 200 future episode judgments. Under an iid worst-case agreement approximation, this gives a 95% screen of about ± 6.9 percentage points. Shared devices, sources and adjudication domains reduce effective sample size; recalculate design effects. Kappa precision requires a prevalence-specific pilot, not this agreement interval.

[PD] Supplement the random core with difficult replans, late-posted receipts, absent scans and disputed entitlements. Such oversampling finds failure modes but needs design weights for population accuracy. Report the full confusion matrix, category-specific errors, agreement and kappa with uncertainty. Reviewers must be blind to the tested workflow’s recommendation when establishing entitlement.

[AI] Begg and Greenes explain why verification limited to cases that receive a definitive check can bias accuracy estimates. Reversals and customer contacts are selected validation channels. Reference linkage or customer response may depend on both the underlying problem and the tested workflow; correction requires credible selection assumptions or explicit bounds.

[PD] Lock the rule and available-input set before testing. Known mandatory linkage failures retain the existing remedy, not a new denial. Independently audit evidence timestamps to prevent hindsight leakage. Reconcile operational access through authorized routes; a failed live query cannot be repaired by inventing a current trend.

Basis: proposed independent classification audit [PD]; Begg and Greenes (1983), Hui and Walter (1980), Cohen (1960). No historical accuracy, agreement or reference completeness is estimated here.

TECHNICAL ANALYSIS

Changes in booked loss do not isolate changes in physical loss

[HR/AI] Available historical comparisons mix bucket definitions, policy periods, source versions and sparse observations. A change in the record can reflect physical events, capture, entitlement rules, financial processing or the observation window.

[AI] The source's slower-growing queue bucket does not establish that an intervention contained it. A one-week movement described as a natural experiment lacks a demonstrated assignment rule, comparator and stable measurement contract. Neither a rising financial count nor a falling contact rate identifies its causal mechanism.

[RC/AI] The private inventory contains 33 source/sensitivity screens, 10 separate version/correction audits and three plan-endpoint audits. None of the 33 crosses tentative limits; no stable phase or eligible repeated Laney panel is established. Preserve vintage and scope alternatives. Individuals limits use $\text{mean} \pm 2.66 \times \text{mean moving range}$; the moving-range upper limit is $3.267 \times \text{mean moving range}$. Sparse points, policy changes, irregular spacing and incompatible exposure make such arithmetic non-diagnostic. A crossing is not treatment attribution; no crossing is not demonstrated stability.

[PD] An operational rate panel requires unique eligible numerators, real exposures, rational subgroups and a stable state definition. Laney adjustment cannot repair identity errors, incompatible denominators or a policy-driven mixture. Do not manufacture weekly counts from rounded rates or use broad reported accuracy as a binomial sample.

[AI] Source governance selects low-performing sites, but an implemented treatment cohort and repeatable untreated trajectories are not established. Under an explicitly assumed linear conditional-mean model, expected untreated follow-up is $m + \rho \times (x - m)$, with population mean m and assumed persistence ρ . Values 0.4, 0.7 and 0.9 imply 60%, 30% and 10% reversion of selected excess. These are sensitivities, not estimated historical fractions.

[PD] Prespecify device/event-capture quality, identity replanning, delivery channel, site type, baseline ambiguity and policy version as moderators. Compare treatment and control on baseline risk and operating load. Do not condition the total effect on captured receipt, reversal, completed return or successful adjudication; these are potential downstream selections.

[AI] Deming's warning about reacting to noise does not excuse a verified arithmetic or entitlement error. Maintain separate processes for statistical escalation, data-contract repair and immediate restoration of an incorrectly withheld customer remedy.

Basis: source-series eligibility assessment [HR/AI], tentative XmR conventions and proposed future process design [PD]. The frozen source-audit inventory governs the recorded screens; no stable phase or historical causal effect is asserted.

TECHNICAL ANALYSIS

Contribution ends at the evidence-supported work product

[HR/AI] The register supports a provisional account of answer construction, qualification of figures and action routing. Its descriptions of author metadata, comments and a build script do not substitute for authentication of those original artifacts.

Contribution claim	Process-tracing grade and boundary
Executive answer and revision	Provisional registry-level hoop: transcribed role, comment and document descriptions support participation; original authorship and complete version history require verification.
Number assurance and open questions	Provisional registry-level hoop: corrections and unresolved pulls are described. Remaining arithmetic conflicts mean assurance did not validate the entire factbase.
Remediation and financial benefit	Unestablished for this case: named fixes and avoidance estimates belong to adjacent programs; no sent-message, accepted funding or mature effect is demonstrated.

[AI] Reviewer challenges, the initial draft, implementation by other teams and the proposed author’s answers are distinct contributions. Agent-generated verification and repeated summaries may share one source. Script output does not alone prove who wrote its logic. The present independent counterexamples and calculations are new analysis, not retroactive credit for the historical reply.

[AI] Identification limits are consequential: scan patterns do not identify custody or prevention; crude channel contrasts do not remove composition and access differences; refund reversal is not a gold standard. The source’s reported contact improvement and accuracy remain unverified comparisons, not causal customer benefits of this work.

[AI] Measurement limits include conflicting lineages, failed aggregate/subgroup alignment, uncertain event timing, partial attribution and disputed implementation status. Replacing an unsupported number with a qualitative claim does not authenticate its direction. A recorded access failure explains missing evidence, not the missing operating result.

[PD] Decision limits remain: the shadow model assumes an adjudicable contract, calibration-quality reference, independent domains and specified costs. Actual ambiguity, harm weights and capacity must be measured. A proposal may be reviewable while deployment, savings and customer benefit remain unestablished; the final case must preserve that distinction.

Basis: derivative role and artifact descriptions [HR]; provisional process-tracing assessment [AI], following Collier. No sole authorship, verified send, implementation ownership or realized savings is claimed.

TECHNICAL ANALYSIS

Test the evidence workflow before changing the refund rule

[PD] The proposed evaluation randomizes independent capture/reconciliation domains to REC or usual evidence processing. It runs in shadow: actual payments, customer notifications, refund eligibility and remedy deadlines remain under the existing approved policy.

Estimand	Fixed population, clock and interpretation
Primary shadow ITT	Usual minus REC expected error loss over every originally eligible episode; incremental setup and coordination are charged once separately.
Customer-protection gate	Usual minus REC all-episode risk of a withholding recommendation for a customer entitled under the original rule. A lower payment rate is not this measure.
Reference endpoint	Independent adjudication by original decision deadline + 28 days reconstructs entitlement at that original deadline; later receipt is separately recorded.
Later field effect	Requires a separately authorized operating intervention, mature customer outcomes and new power. The shadow result cannot estimate realized cash or customer harm avoided.

[PD] Freeze eligibility before new capture or linkage, retain false custody flags, missing responses and known blockers, and use the same approved information rights in both arms. Domain boundaries must encompass shared devices, support queues and reconciliation staff. Distinct frozen batches are sampled from the same frame; they are not assumed to be paired realized episodes.

[PD] Lock the scoring rule and input timestamps before reference outcomes are available. Known evidence failure preserves the live remedy and the protocol-defined shadow fallback. The numerical model’s fully specified probabilities are hypothetical; reference uncertainty, blocked-case frequency, residual dependence and actual error costs require a pilot.

[AI] Power must resolve a benefit at the fixed implementation cost and clear the wrongful-withholding gate. Strong power at a large assumed benefit does not establish that economic break-even is detectable. The model pages test that constraint under declared finite laws; their calibration does not certify field prevalence or practical recruitment.

[PD] A later staggered operational adoption could target group-time ATT against not-yet-treated domains, with no anticipation, overlap, credible conditional parallel trends and interference checks. Prespecify customer waiting, unresolved entitlement and dispute outcomes as well as finance. Joint pretrend diagnostics and equivalence do not replace identification. This extension remains unpowered and unestimated.

Basis: proposed randomized shadow benchmark and separate field extension [PD]; Callaway–Sant’Anna (2021) applies only to the latter. No withholding, refund delay or customer action is executed by this case.

TECHNICAL ANALYSIS

Four tests keep evidence improvement separate from remedy suppression

[AI/PD] REC is a proposed framework. Its purpose is to improve justified decisions while keeping physical truth, evidence quality, contractual rights and financial records distinct. It does not certify an owner or a policy from a convenient label.

Proposition	Prespecified observation that challenges it
P1: Better links improve decision evidence	Independent deadline-matched auditing finds no material gain in valid evidence, or added false links offset the gain.
P2: Reconciliation protects entitled customers	Wrong-withholding recommendations fail the required improvement or worsen in a prespecified channel or low-capture stratum.
P3: The workflow justifies its cost	The all-episode benefit bound does not clear incremental cost after reference-error and adverse-cost sensitivity.
P4: Ownership routes to an actionable control	A proposed owner lacks the authority or feasible action specified in the contract; a controlled test does not support the claimed preventive pathway.

[AI] Diagnostic literature explains why agreement between dependent records is insufficient and why selective verification can conceal errors. Event-history methods explain why later receipt and earlier entitlement need separate state definitions. Vickers and Elkin connect model use to asymmetric decision consequences; they do not supply this case’s valuations or authorization.

[AI] Smith, Bolton and Wagner study customer evaluations of service-failure recovery in scenario experiments. Their work motivates measuring the remedy experienced by the customer separately from the provider’s accounting result; it does not transfer a numerical parcel-service effect. The source’s lower reported contact rate is therefore informative context, not a sufficient welfare endpoint.

[AI] Campbell and Goodhart warn against target-driven distortion; an incentive to lower recorded refunds can reward silent under-remedy. Ittner and colleagues caution that additional measures or claimed alignment do not automatically improve performance. OECD composite guidance supports transparent weights, but an entitlement or authority failure must remain a noncompensating gate.

[PD] Register material effects, subgroup protections, reference procedures and stopping rules before evaluation. Preserve negative findings and unresolved truth. An imprecise test is inconclusive; a better financial metric alone is insufficient. Route an evidenced capture defect to its owner while retaining the customer’s existing remedy until a separately authorized policy decision is justified.

Basis: proposed synthesis [AI/PD] grounded in newly checked primary diagnostic, event-history, decision-analysis and service-recovery sources; continuing incentive literature. No cited paper validates REC or supplies a historical treatment effect.

A lower payment recommendation rate can hide a worse remedy decision

[PD] Randomize independent event-integration domains to receipt-evidence capture plus linkage reconciliation or usual capture. This is a shadow decision benchmark: live payments, approved entitlement rules and customer deadlines remain unchanged.

[PD] Freeze 80 original ambiguous-receipt episodes per domain before assignment. Both arms have equal access and authority. Independent batches share one eligibility law; latent easy/hard domain difficulty correlates episodes. A blinded reference by original decision deadline + 28 days reconstructs entitlement at that deadline. Later delivery cannot rewrite that earlier truth.

Shadow workflow	Recommend pay	Wrong withholding	Resource loss
Usual capture	71.7%	2.6%	0.4501
Capture only	56.2%	4.7%	0.4431
Capture + reconcile	60.1%	0.9%	0.2136

[PD] Purported qualifying evidence may be false. Known failed or unresolved mandatory linkage, modeled for 10% of original episodes, preserves the existing remedy in both arms. It creates no new withholding recommendation. Entitlement is independently judged under the approved contract, not inferred from physical receipt alone.

DO NOT OPTIMIZE PAYMENT REDUCTION

[PD] Capture-only is a separate counterfactual, not a randomized third arm. More records can reduce payment recommendations while increasing wrongful-withholding recommendations. No payment is actually withheld in this benchmark.

[PD] Entirely invented probabilities and resource prices. All original episodes remain; the source accuracy and reversal summaries do not calibrate this law.

Require decision value and fewer wrongful-withholding recommendations

[PD] Finite domain-score convolutions preserve shared difficulty and independent domains. Increasing response strength improves capture and specificity on a declared parameter path. Exact rational CDF checks establish marginal ordering on that path.

Design	Per arm	MDE (index)	BE detection	Joint floor
Initial	8	40.79	29.6%	0.0%
Base candidate	32	20.84	85.1%	90.6%
Same size / stress	32	21.01	84.4%	4.8%
Conditional redesign	112	11.26	>99.9%	84.1%

[PD] Break-even is 22.22 index points. Require 80% correctly directed detection at break-even and 80% joint power for an economic lower bound above cost and a reduction in wrongful-withholding recommendations. Each decision gate has a 2.5% one-sided error bound. The separate two-sided zero test has at most 5% error.

[PD] Stress retains half the assumed specificity improvement while preserving capture gains. A union bound gives the joint-power floor; it is not an assumed independence product. MDE and confidence inversion use the actual finite law at the candidate effect, not a constant full-effect variance.

[PD] Retaining only one-quarter of specificity improvement gives a joint floor of 20.4% at the redesign. A zero floor would be uninformative, not zero actual joint success. Reference error, shared reviewers and unknown operational frequencies require fresh calibration.

CONDITIONAL RECRUITMENT

[PD] The redesign requires 224 independent domains and 17,920 original episodes. Their availability is unverified. No deployment, customer safety or cash benefit is established by this sizing.

[PD] Fixed usual shadow resource loss = 100. Known-law discrete bounds are conditional planning results; model correctness and reference validity are not established by arithmetic.

Count entitled remedies as obligations, never as avoided cost

[PD] An unnecessary payment costs one arbitrary resource unit; a wrongful-withholding recommendation costs six decision-risk units. Correct recommendations cost zero. These are shadow loss weights, not observed cash or an assertion that customer harm can be traded away.

[PD] New work $K = 0.10$ per original episode includes capture 0.040, linkage verification 0.035 and support 0.025, outside error loss. $Net_raw = a \times h \times 0.133875 + 6 \times 0.017100 - K$. Avoidability a and non-overlap h apply only to confirmed unowed-payment value; owed payments are never savings.

Avoidability a	$h = 0.50$	$h = 0.75$	$h = 1.00$
0.50	8.01	11.73	15.45
0.75	11.73	17.31	22.89
1.00	15.45	22.89	30.32

[PD] Table entries are net index points at $K = 0.10$; divide raw net by 0.4501 and multiply by 100. At $a = h = 1$, $K = 0.075 / 0.100 / 0.150$ gives 35.88 / 30.32 / 19.21 points. Reprice and re-power changed scenarios.

[PD] If each arm has at most ϵ missing or misclassified reference truth, worst-case loss-contrast uncertainty is 12ϵ and safety uncertainty is 2ϵ . One-percent uncertainty already removes the positive safety lower bound. Retain every unresolved episode; a day-28 reference is not automatically ultimate truth.

[PD] A future field net would price causal reductions in confirmed unowed payments, debit added wrongful withholding, waiting and work, and subtract deployment cost over the same original cohort. Those component effects are unestimated and unpowered here.

KEEP TWO DECISIONS SEPARATE

[PD] Shadow value can justify better evidence collection. Changing payment or reversal policy requires separate authority and independently validated remedy protections.

[PD] Research-only hidden adjudication is a separate study expense. Positive weighted value cannot override the independently powered wrongful-withholding gate or authority requirements.

PROPOSED INFORMATION VALUE / 4 OF 4

Buy evidence that can redirect investment, not justify a predetermined restriction

[PD] A separate three-state model prioritizes entitlement-rule mismatch, including concurrent capture gaps; otherwise it distinguishes capture/linkage failure from neither material gap. This partition guides evidence investment, not permission to change remedies.

Action / state	Capture gap	Rule mismatch	Current correct
Capture validation	14	-12	-7
Rule reconciliation	2	11	-3
Retain workflow	1	2	5

[PD] Priors are 0.40 / 0.25 / 0.35. Signal accuracy q is the probability of the true state, with each wrong state receiving (1 – q)/2. Downstream action expense is already in the payoff table; information acquisition is charged separately.

Accuracy	Gross EVSI	Acquisition	Net EVSI
0.50	1.2000	0.50	0.7000
0.75	4.3250	1.25	3.0750
0.90	6.2000	2.50	3.7000

[PD] All 27 contingent policies are enumerated. Prior optimal value is 2.6500; EVPI is 7.4500. Gross EVSI cannot exceed EVPI. Utilities are arbitrary and are never added to modeled error reduction, refund exposure or historical savings.

PRESERVE THE CURRENT REMEDY

[PD] A weak or missing signal can support retaining the approved workflow. A stronger signal can support bounded capture validation or rule clarification, while the same access, entitlement and customer-protection gates remain mandatory.

[PD] All priors, signal accuracies and utilities are hypothetical. The information decision does not assume that capture investment or policy tightening has a positive effect.

Reconstruct the decision that was due, using evidence available at that time

[AI] The Remedy Evidence Contract distinguishes physical custody, recorded events, authorized entitlement and accountable action. Its proposed value lies in better evidence and decisions; lower payments alone do not define success.

Claim class	Meaning and replication boundary
[HR] Historical record	The supplied register describes the executive question, rewritten answer, corrections, ownership disputes and unresolved requests. It does not establish that the communication was sent or that any proposed fix was implemented.
[RC] Recalculated	Endpoint arithmetic, implied exposures, contribution residuals, pooled-rate feasibility, tentative process screens and observation counterexamples. Public numerical constructions are invented; source values and version conflicts remain private.
[AI] Analytical inference	Missing events do not establish physical custody. Reversal frequency does not identify original accuracy. A lower payment total cannot distinguish evidence improvement from wrongful withholding without an independent entitlement reference.
[PD] Proposed design	An evidence contract and shadow evaluation that preserve live remedy authority and timing. All evaluation probabilities, resource weights, intervention costs and information values are planning assumptions.

Replication sequence

[PD] Freeze the register and recover original drafts, comments, answer files, code, readouts and deployment records. Authenticate the contribution chain before upgrading its process-tracing grade. Preserve the difference between the document prepared and the communication actually sent.

[PD] Recompute each metric within its population, definition, vintage and maturity window. Verify the channel-weight feasibility test, complete signed bridges and all source alternatives. Keep inferred denominators distinct from observed counts; do not repair incompatible branches by averaging them.

[PD] Reproduce the observation counterexample and evaluation law. A later reference may reconstruct original entitlement; it must not let later receipt rewrite what the earlier decision was required to do. Bind numerical, source and visual reviews to the released files.

THE INTERPRETATION THIS CASE CHANGES

[AI] The missing control is not simply a stricter refund rule or a more forceful owner label. It is a verified link from the original operating event to an authorized remedy and an accountable next action.

The separate private package contains raw values, identity mappings, source locators and executable checks. The public case discloses methods and protected conclusions.

RESEARCH REFERENCES / OBSERVATION AND CUSTOMER REMEDY

Validate the state behind the record

Primary research supports the interpretation and proposed evaluation. It does not authenticate unrecovered project artifacts or provide a historical effect, deployment result or financial return.

Hui, S. L., & Walter, S. D. (1980). Estimating the Error Rates of Diagnostic Tests. *Biometrics*, 36(1), 167–171. DOI 10.2307/2530508. Original article. Error rates without a perfect reference require identifying assumptions, including distinct prevalences, invariant test errors and conditional independence in the two-test model. Shared delivery evidence can violate independence; a reversal stream is not an independent gold standard.

Begg, C. B., & Greenes, R. A. (1983). Assessment of Diagnostic Tests When Disease Verification Is Subject to Selection Bias. *Biometrics*, 39(1), 207–215. PMID 6871349. Primary abstract. Selective verification can bias accuracy estimates. Audit the original eligible population, including unreversed and unresolved cases; any correction for missing verification needs a defensible observation mechanism.

Vickers, A. J., & Elkin, E. B. (2006). Decision Curve Analysis: A Novel Method for Evaluating Prediction Models. *Medical Decision Making*, 26(6), 565–574. DOI 10.1177/0272989X06295361. Original author manuscript. Decision usefulness depends on error consequences and threshold preferences, not accuracy alone. This case adopts explicit error costs and a separate customer safeguard; no clinical effect or threshold is transferred.

Aalen, O. O., & Johansen, S. (1978). An Empirical Transition Matrix for Non-Homogeneous Markov Chains Based on Censored Observations. *Scandinavian Journal of Statistics*, 5(3), 141–150. Original article, selected model and censoring sections. Distinguish time-indexed transitions and observation. A later delivery or reversal does not automatically reveal the earlier state. No transition estimator is fitted without the required longitudinal event records and assumptions.

Hernán, M. A., Hernández-Díaz, S., & Robins, J. M. (2004). A Structural Approach to Selection Bias. *Epidemiology*, 15(5), 615–625. DOI 10.1097/01.ede.0000135174.63482.43. Original article reproduced on PDF pages 20–30 of the linked teaching collection. Conditioning on an observed event or selected review can open noncausal paths. Freeze original eligibility rather than comparing only the cases a new capture process makes visible.

Smith, A. K., Bolton, R. N., & Wagner, J. (1999). A Model of Customer Satisfaction with Service Encounters Involving Failure and Recovery. *Journal of Marketing Research*, 36(3), 356–372. DOI 10.1177/002224379903600305. Publisher abstract only. Scenario experiments in restaurants and hotels examine matching recovery to failure. The conceptual application is to evaluate remedy quality alongside service failure; logistics effect sizes, customer utility and causal claims are not imported.

Access scope is recorded privately. Literature applications are methodological or conceptual; outside empirical effects are not attributed to this project.

RESEARCH REFERENCES / EVALUATION AND UNCERTAINTY

Retain the original population and its dependencies

Primary research supports the interpretation and proposed evaluation. It does not authenticate unrecovered project artifacts or provide a historical effect, deployment result or financial return.

NIST/SEMATECH. e-Handbook of Statistical Methods, section 6.3.2.2: Individuals Control Charts. Official method. Tentative XmR bands are mechanical diagnostics when observations are few, inferred or differently scoped. Preserve comparable source branches and distinguish a crossing from an identified treatment effect. A short within-limit movement does not establish improvement.

Laney, D. B. (2002). Improved Control Charts for Attributes. *Quality Engineering*, 14(4), 531–537. Primary publisher abstract. Repeated subgroup counts and exposures are needed for an eligible attribute panel. A reported accuracy percentage, a reversal range and isolated approximate counts do not supply a valid large-sample rate series.

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20, 37–46. Primary publisher record. Independently classify receipt evidence, entitlement and ownership. Report agreement, the confusion matrix, prevalence and uncertainty; agreement on a shared erroneous source is not validation.

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. Primary abstract. Extreme-unit selection and peak-period comparisons can generate apparent recovery without treatment benefit. Keep any reliability-based reversion scenario explicitly assumed when repeated untreated data are unavailable.

Hudgens, M. G., & Halloran, M. E. (2008). Toward Causal Inference With Interference. *Journal of the American Statistical Association*, 103, 832–842. Original manuscript. Shared event infrastructure, devices, reviewers or operational teams can couple episodes. Define independent domains and assess spillovers before treating recommendations as independent observations.

Callaway, B., & Sant'Anna, P. H. C. (2021). Difference-in-Differences with Multiple Time Periods. *Journal of Econometrics*, 225, 200–230. Author manuscript. A later operational rollout could estimate group-time effects under explicit comparison, timing, overlap and parallel-trend assumptions. The proposed shadow evaluation does not estimate or power that separate customer-outcome study.

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. Author repository. Value information by the actions its signal changes, within a declared utility model and after acquisition cost. Do not add evidence-purchase utility to the same operating benefit it may help realize.

Access scope is recorded privately. Literature applications are methodological or conceptual; outside empirical effects are not attributed to this project.

RESEARCH REFERENCES / CONTRIBUTION AND GOVERNANCE

Keep measurement, accountability and remedy separate

Primary research supports the interpretation and proposed evaluation. It does not authenticate unrecovered project artifacts or provide a historical effect, deployment result or financial return.

Collier, D. (2011). Understanding Process Tracing. *PS: Political Science & Politics*, 44, 823–830. Primary publication. Test specific claims about answer authorship, revision, acceptance and execution. Derivative descriptions of a byline or script remain provisional until the original is authenticated; no amount of repetition establishes a realized program benefit.

Tucker, A. L., & Edmondson, A. C. (2003). Why Hospitals Don't Learn from Failures. *California Management Review*, 45(2), 55–72. Author manuscript. Repeated exceptions can reveal missing operating capabilities rather than isolated worker errors. The proposed evidence contract makes observation and ownership gaps visible; its effectiveness still needs evaluation.

The W. Edwards Deming Institute. The Funnel Experiment. Official exposition. Repeated adjustment to routine variation can worsen performance. A short-period defect change should not by itself trigger a new owner verdict or tighter customer-remedy rule.

Campbell, D. T. (1976). Assessing the Impact of Planned Social Change. Original institutional record. Targeted measures can alter behavior and the recording process. Preserve original eligibility and measure wrongful withholding, unresolved evidence and customer burden alongside recorded losses and payment totals.

Goodhart, C. A. E. Problems of Monetary Management: The U.K. Experience. 1975 conference; revised volume listed as 1976 in the RBA archive. A targeted proxy can lose its earlier meaning. Refund reduction and lower observed loss are therefore separated from correct entitlement and physical service; neither financial direction alone establishes success.

Ittner, C. D., Larcker, D. F., & Randall, T. (2003). Performance Implications of Strategic Performance Measurement in Financial Services Firms. *Accounting, Organizations and Society*, 28, 715–741. Author manuscript. Connect measures to the operating mechanism and decision rather than equating more indicators with better performance. An ownership table, score or completed reply does not establish improved customer outcomes.

OECD / European Commission JRC (2008). Handbook on Constructing Composite Indicators. Official handbook. Aggregation requires defensible definitions, weights and assumptions about compensation between dimensions. The proposed case keeps customer-remedy safeguards separate from economic utility; a favorable weighted average cannot authorize a failed entitlement condition.

Access scope is recorded privately. Literature applications are methodological or conceptual; outside empirical effects are not attributed to this project.

FINAL REFEREE REPORT / INTERNAL AGENT-ASSISTED REVIEW

Accept the diagnostic contribution; test the remedy safeguards

[AI] This author-led, agent-assisted review is not external peer review or client approval. Its disposition concerns evidence reconciliation, the new identification findings and a proposed shadow evaluation.

Major issue	Disposition in the revised case
Custody inferred from a missing event	Separated physical state, observation, decision-time knowledge and entitlement. Historical owner verdicts are reported as interpretations; causal attribution remains conditional on validated custody evidence.
Reversal rate treated as accuracy	Constructed observationally equivalent worlds with different original states and capture quality. A later delivery can change the state without proving the original refund was wrong.
Scope used to excuse arithmetic	Recalculated within-lineage comparisons and tested pooled/channel feasibility. Preserved source versions and complete signed-bridge requirements instead of selecting the most favorable headline.
Recorded correction treated as certification	Retained unresolved source conflicts and distinguished missing instrumentation from unavailable retrieval. A deterministic rewrite does not authenticate its inputs or demonstrate final acceptance.
Payment reduction treated as value	Specified shadow-only comparison with original eligibility, independent entitlement reference, explicit error costs and a separately powered harmful-recommendation gate. Owed remedies remain protected.

Partner critique: can a COO act on page one?

[AI] Yes: require a compatible factbase, assign the next evidence and action owners, and protect existing remedies while testing event capture and reconciliation. The immediate decision funds measurement and a bounded shadow test; it does not authorize unverified refund restrictions or claim ownership of adjacent savings.

Remaining conditions

[AI] Recover originals; resolve metric scope; validate custody, entitlement and timing; calibrate domain dependence, reference error, costs and feasible recruitment. Sending, acceptance, execution and mature customer impact remain unestablished.

REVIEW DISPOSITION / PASS WITHIN STATED SCOPE

[AI] The diagnostic case and proposed REC evaluation pass the stated internal review gates. This verdict does not authenticate unrecovered originals, establish historical causal impact or certify customer-facing deployment.

Internal evidence, methods, numerical and executive review. Release acceptance is limited to the documented source, model and confidentiality boundaries; no outside endorsement or ranking is implied.