

Chain-of-Custody Diagnostics

Turn a disputed return diagnosis into an accountable operating test

An anonymized operating case on returned shipments: reconciling conflicting driver-identity measures, testing a loading-error explanation, and designing programmatic actions without mistaking a scorecard label for causal responsibility.

PROJECT CONTEXT / LOGISTICS QUALITY, PARTNER GOVERNANCE & DIAGNOSTIC INTEGRITY

BLUF / BOTTOM LINE UP FRONT

Trace the handoff before assigning blame.
Test the control before claiming benefit.

ANALYTICAL FINDING

The same return-identity and sort-compliance table can arise from a loading error or a legitimate later transfer. Stable percentages do not resolve that ambiguity. Verify the custody sequence and the population before selecting a corrective intervention.

CONTRIBUTION IN THE SOURCE RECORD

Provisional: directed the initial analysis and action framing. A teammate authored the replacement; adoption remains unverified.

EVIDENCE STATUS

Register-based diagnostic and operating design; originals unrecovered. Causal responsibility, adoption and financial benefit remain unestablished.

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

STRUCTURED ABSTRACT / THE EXECUTIVE DECISION

Approve custody verification before a broad coaching response

[AI] The business question was practical: which recurring controls should reduce undelivered shipments? Answering it requires more than a return code. The identity being compared, the events selected and the mechanism causing failure must all match the proposed action.

Element	Decision-relevant finding
Context [HR]	An executive review requested a root-cause Pareto and partner actions for shipments reported missing on a route and later returned. The initial framing favored driver coaching; the analysis challenged that assumption.
Objective [HR]	Identify an actionable split, test whether a targeted coaching population existed, and route controls through established station and partner mechanisms. The assignment was analytical; implementation was not demonstrated.
Approach [HR/RC]	The work compared return identity with sort status, reconciled grain and denominator choices, examined event concentration, and screened alternative data sources. A separately authored replacement changed the measurement basis and priorities.
Findings [RC/AI]	A stable label ratio does not establish physical cause. Under compatible reported summaries, most workers had one event but most events involved repeated-worker records. Neither that pattern nor return identity establishes coachability.
Decision value [PD]	Approve an independently adjudicated custody sample and a bounded control evaluation. Require reconciled identity roles, common eligibility, frozen event time and a confirmed causal mechanism before a finding changes partner accountability.
Implications [PD]	The data owner reconciles definitions; station and partner leads investigate their mechanisms; the sponsor approves the intervention. Finance values mature service effects after incremental costs and overlap, rather than treating recorded compensation as savings.

THE OPERATING DECISION

[PD] Apply the Custody-to-Action Protocol (CAP): validate the population, reconstruct the handoff, select the owner-controlled intervention, then test mature service and accountability outcomes. Keep unverified events out of punitive individual rankings.

[HR] Source S1: selected project register and its versioned evidence pointers. [RC] recalculated; [AI] analytical inference; [PD] proposed design. Source identities and raw values remain private.

EXHIBIT 1 / ASK WHICH IDENTITIES ARE BEING COMPARED

Two measures named “same driver” can answer different questions

[HR] One source compared the exception marker with the recorded returning worker, combining a different worker and no recorded worker in one class. Another compared the assigned worker with the returning worker. Their populations and rules differed.

ONE SHIPMENT / THREE DIFFERENT IDENTITY ROLES

Constructive example: assigned A, exception marker B, returning A

ASSIGNED
A
Planned responsibility

EXCEPTION MARKER
B
Recorded observation

RETURNING
A
Receipt association

Assigned = returning; exception marker ≠ returning.
Both comparisons can be correct. Neither identifies a physical failure or preventable cause.

Separate different identity from unknown identity; preserve the comparison, clock, null rule and assignment version.

Identity role	What its agreement or disagreement can establish
Assigned worker	A planning or routing association at a specified version and time. Assignment does not by itself establish physical possession.
Exception-marking worker	The recorded actor who marked the route exception. Validate actual actor, account use and whether the event represents direct observation.
Returning worker	The actor linked to station receipt. Authorized rescue, transfer, shared debrief or missing attribution can separate this actor from earlier custodians.

[AI] If assigned and returning workers match while the exception marker differs, both comparisons can be correct and yield opposite classifications. A mixed different-or-unknown class further prevents counting confirmed mismatches without a null decomposition. Neither comparison proves cause; different populations add a selection problem.

REQUIRED METRIC CONTRACT

[PD] Publish the exact pair of identity roles, null handling, grain, eligibility, clock and assignment version. Reserve “same driver” as a short label only after those fields are fixed.

[HR] Source S1: selected project register and its versioned evidence pointers. [RC] recalculated; [AI] analytical inference; [PD] proposed design. Source identities and raw values remain private. Identity diagram is a proposed measurement contract.

EXHIBIT 2 / SEPARATE A MECHANISM HYPOTHESIS FROM A FINDING

A return mismatch does not identify where physical custody failed

[HR] The initial diagnosis treated differently attributed returns with compliant sorting as evidence favoring station loading. The action set included custody reconciliation and cart-to-route verification; those proposals had not confirmed the mechanism.

SAME SUMMARY / DIFFERENT CUSTODY HISTORIES

Constructive examples: compliant sort and different exception-marker / return identities

HISTORY A
Correct sort → wrong vehicle at loading
Later return by another worker
A loading defect is present

HISTORY B
Correct sort → correct vehicle at loading
Authorized transfer → another worker returns
No loading defect is required

The observed table cannot distinguish these histories.
Trace the time-ordered handoffs before assigning a causal owner.

A plausible mechanism is a reason to investigate; the proposed trace is not evidence that the trace already occurred.

[AI] Consider two histories with the same observed identity mismatch and compliant-sort flag. In one, the shipment enters the wrong vehicle at loading. In the other, loading is correct and a later authorized transfer changes the returning worker. The summary table cannot distinguish them. A recorded compliant state may itself require validation.

Inference that survives	Inference requiring additional evidence
Investigate the loading interface	The table can nominate a plausible inspection point. It cannot identify the fraction caused by that interface.
Examine shared station conditions	Several partners at one station can justify a station audit. Shared workload, routes, recovery practices or coding can also produce their common pattern.
Protect individual attribution	Without custody validation and exposure, event counts are insufficient grounds for naming or penalizing a worker.

DECISION-CHANGING RESULT
[AI] The proposed loading control remains testable, but the evidence supports an investigation owner rather than a proven causal owner. A stable driver-match ratio does not close this identification gap.

[HR] Source S1: selected project register and its versioned evidence pointers. [RC] recalculated; [AI] analytical inference; [PD] proposed design. Source identities and raw values remain private. Alternative histories are constructive analytical examples, not claims about what occurred.

EXHIBIT 3 / RECONCILE THE POPULATION BEFORE SCALING THE ANSWER

A precise selected-panel percentage can misstate the network opportunity

[HR] The first panel selected a narrow slice of the broader defect population, with no-return records outside it. A second flag-based panel had limited overlap and different inclusion rules. The replacement used another source and revised the priority.

THREE TESTS / EACH ANSWERS A DIFFERENT QUESTION

Conceptual validation sequence; no historical panels are pooled

ARITHMETIC
Does the result survive grain and tie handling?

MEASUREMENT
Does the label represent the intended event?

TRANSPORT
Does this selected panel answer the wider decision?

A small change under duplicate handling can coexist with a large interpretation error.

No-return records are not an untreated control or automatically missing eligible return data.

An inclusion rule, join scope, version and maturity window must travel with every percentage.

Validation question	Consequence for the decision
Does the filter select an outcome?	A returns-only denominator describes a selected outcome. Its small share of all defects is not automatically missing attribution among eligible returns. Row partitions also need a distinct-entity overlap check.
Does the join recover the intended volume?	A high key-match rate can join a special-value subset or a defect-specific table. Validate scope and expected volume per worker before calculating a rate.
Do alternate panels overlap representatively?	A small and unusually affected overlap does not validate transport to the unlinked population. Compare inclusion, taxonomy, time and maturity separately.
Is the replacement a like-for-like replication?	A new source, different period and revised classification can corroborate a question while failing to reproduce its original quantity. Preserve both versions.

[AI] Three different tests are needed: whether the arithmetic is stable, whether the metric measures the intended construct, and whether the measured population supports the intended decision. Passing duplicate sensitivity addresses the first test. It cannot substitute for the other two.

[HR] Source S1: selected project register and its versioned evidence pointers. [RC] recalculated; [AI] analytical inference; [PD] proposed design. Source identities and raw values remain private. No source panels are pooled or treated as interchangeable.

EXHIBIT 4 / TARGET THE MECHANISM, NOT THE LARGEST WORKER COUNT

Most workers having one event does not mean most events are isolated

[RC] Under the reported common-cohort summaries, most affected workers had one event, yet most events were associated with workers who had repeated events. The source’s broad claim that no repeat population existed does not follow from its own totals.

A MAJORITY OF PEOPLE IS NOT NECESSARILY A MAJORITY OF EVENTS

Analytical distinction / then test the mechanism and valid exposure



Single-event workers can be the majority while repeat-worker records carry most events.
Neither population share nor event share establishes fault, preventability or treatment response.

Freeze the list and cutoff before follow-up; account for selection, shared conditions and regression to the mean.

[RC/AI] Subtract singleton workers from total events to recover events associated with repeated-worker records, provided the summaries share a cohort and counting rule. The resulting majority does not prove misconduct or coachability. A chosen list’s event share bounds its immediate coverage under perfect removal; it does not estimate preventability or future response.

Screen	Appropriate interpretation
Event concentration	Measure how much observed volume a specified cutoff reaches. Keep the cutoff with the result; changing the list changes the coverage.
Exposure-adjusted recurrence	Requires eligible delivered volume, stable assignment and repeated periods. A special-purpose denominator cannot establish worker risk.
Provider or station over-index	Use uncertainty to prioritize investigation. A lower confidence bound does not remove selection, shared conditions or multiple-comparison risk.
Selected worst performers	Freeze selection before follow-up and compare with suitable untreated units. Apparent improvement can reflect regression to the mean.

RETAIN THE CONTRIBUTION; CORRECT THE INFERENCE

[AI] Blocking unsupported individual rankings was defensible. Mostly single-event workers do not imply mostly isolated events; repeated records still require valid exposure, grain reconciliation and independent mechanism evidence.

[HR] Source S1: selected project register and its versioned evidence pointers. [RC] recalculated; [AI] analytical inference; [PD] proposed design. Source identities and raw values remain private. Repeat-volume arithmetic is conditional on the reported common cohort; no worker-level causal response is estimated.

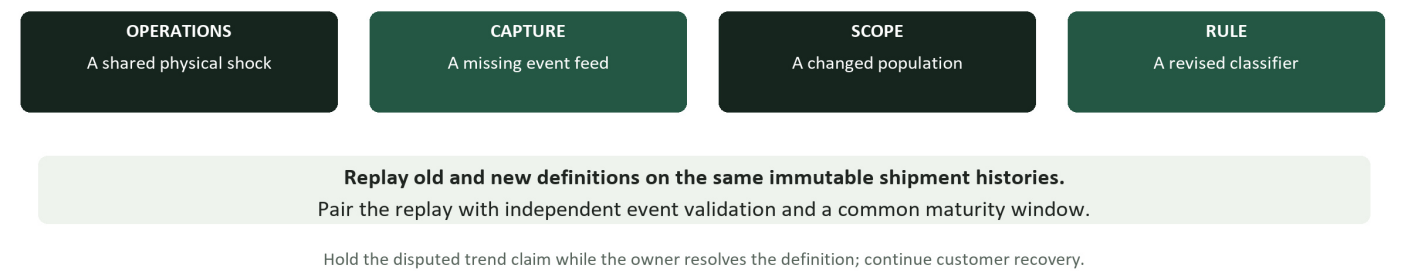
EXHIBIT 5 / TREAT SYNCHRONIZED CHANGES AS A MEASUREMENT QUESTION

A coordinated category jump requires a definition check before a performance judgment

[HR] Several non-containerized categories increased together in a weekly export, while the total defect count moved much less. The replacement author requested confirmation of the categorization logic and recommended holding comparisons spanning the suspected change.

A SHARED CATEGORY JUMP HAS SEVERAL POSSIBLE ORIGINS

Analytical alternatives, not a historical finding of cause



[AI] A shared jump is consistent with a measurement change, but it is not proof of one. A common operating shock, changed scope, missing event feed or new classification rule can all move multiple categories together. The decisive evidence is a versioned rule and replay on a fixed shipment cohort, supported by physical-event validation.

Test	Evidence required to release the hold
Definition replay	Run old and new rules on the same immutable events; distinguish reclassification from changed underlying operations.
Event-capture audit	Trace a stratified sample independently of the category flag. Missing scan evidence must not automatically mean missing physical work.
Process comparison	Assess tentative limits and phase stability. Sparse observations cannot establish a durable operating shift or causal effect.
Maturity check	Freeze a common observation lag; preserve later revisions. A classification settling later is not evidence of service recovery.

[PD] The dashboard owner owns the definition decision. The operating lead owns service containment. Hold the disputed trend claim while continuing customer recovery and evidence collection.

[HR] Source S1: selected project register and its versioned evidence pointers. [RC] recalculated; [AI] analytical inference; [PD] proposed design. Source identities and raw values remain private. The rule change remains a hypothesis until the responsible owner and replay evidence confirm it.

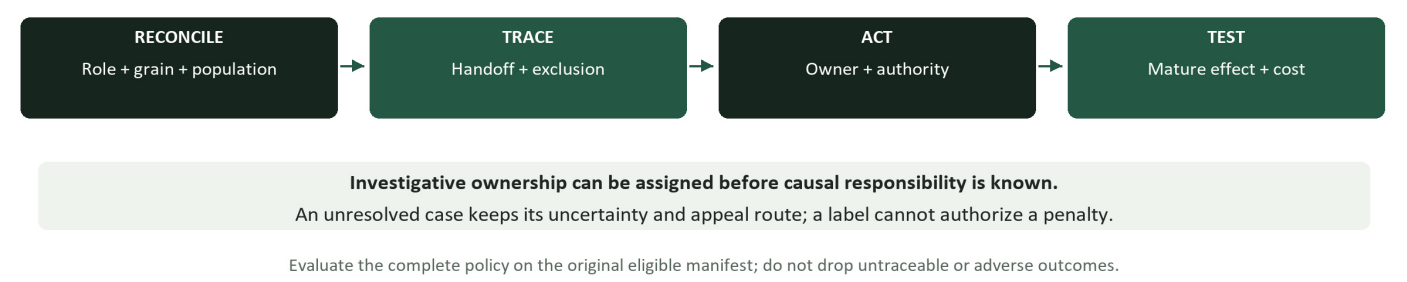
EXHIBIT 6 / MAKE THE ACTION SET EXECUTABLE

Named owners and measurable controls are separate requirements

[HR] The initial action list used existing station and partner channels and included exclusions for events outside the worker’s control. A separate companion assessment found uneven ability to measure its actions. Those were different lists, not successive totals for one program.

CUSTODY-TO-ACTION PROTOCOL / PROPOSED

A control becomes executable when authority, evidence and outcome agree



Action family	Control contract
Custody reconciliation	Station lead traces expected and actual handoffs, including authorized transfers. Define a recovery decision and a reopening rule.
Partner routine	Partner-facing lead tests search, loading or debrief standard work against its specific mechanism. Preserve safety and authorized exceptions.
Reason and identity capture	Data owner measures completeness, correctness and delay separately. Validate the reference independently of the system being improved.
Attribution correction	Policy owner confirms the scorecard definition and correction authority. Record joint responsibility when more than one mechanism contributes.
Exclusion and appeals	Keep an auditable reason for exclusions and disputed cases. Publish eligible coverage so apparent improvement cannot come from removing difficult events.

RELEASE RULE

[PD] An action proceeds when its owner, authority, mechanism, denominator, outcome, guardrail and evidence source are explicit. An owner-filled row alone does not demonstrate adoption or performance.

[HR] Source S1: selected project register and its versioned evidence pointers. [RC] recalculated; [AI] analytical inference; [PD] proposed design. Source identities and raw values remain private. CAP and the release rule are proposed; historical adoption is not established.

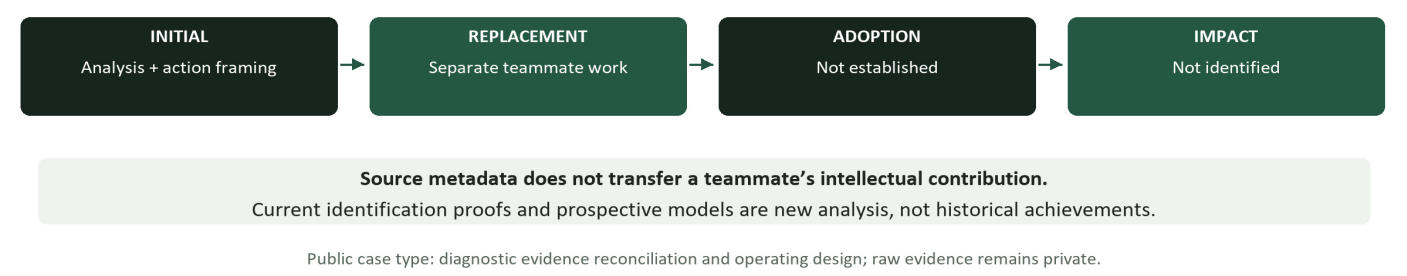
EXHIBIT 7 / ATTRIBUTE THE ANALYTICAL CONTRIBUTION PRECISELY

The work challenged the initial response; it did not establish a realized transformation

[HR] The register distinguishes the initial analytical owner’s judgments from a teammate’s independent follow-up and replacement note. Document metadata and shared storage do not resolve that distinction on their own.

A BOUNDED ANALYTICAL CONTRIBUTION / NO VERIFIED IMPACT

The initial response, replacement analysis and downstream execution have different owners



Claim and process-tracing grade	Supported boundary
Initial analysis and action framing / provisional hoop	The derivative record links the analytical owner to the assignment, population challenges, stress tests and rejection of unsupported individual rankings. Original provenance governs any stronger grade.
Replacement note / separately attributed	The teammate’s first-person content and recorded follow-up support a distinct analytical contribution. An owner name in file metadata does not transfer authorship.
Recommendation adoption / unestablished	A reported deliverable and populated owner list do not show that the sponsor approved, funded or executed the actions.
Mature service and financial benefit / unestablished	No comparison identifies an intervention effect. Historical compensation paid is cost incurred, not an avoidable benefit attributable to this work.

[AI] The transferable contribution is disciplined response design: question the denominator, distinguish identity from cause, stop unsupported individual attribution, and make the next investigation answerable. This case adds a custody-identification proof and a prospective evaluation; those are current analytical work.

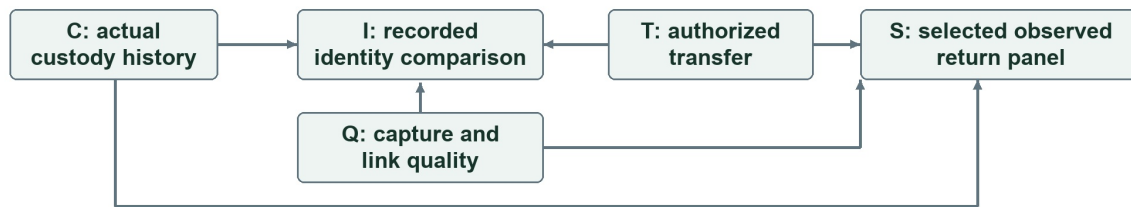
[HR] Source S1: selected project register and its versioned evidence pointers. [RC] recalculated; [AI] analytical inference; [PD] proposed design. Source identities and raw values remain private. Contribution grades apply to the stated stage; no sole-authorship, deployment or causal-benefit claim is made.

TECHNICAL ANALYSIS

Trace a handoff before choosing the preventive control

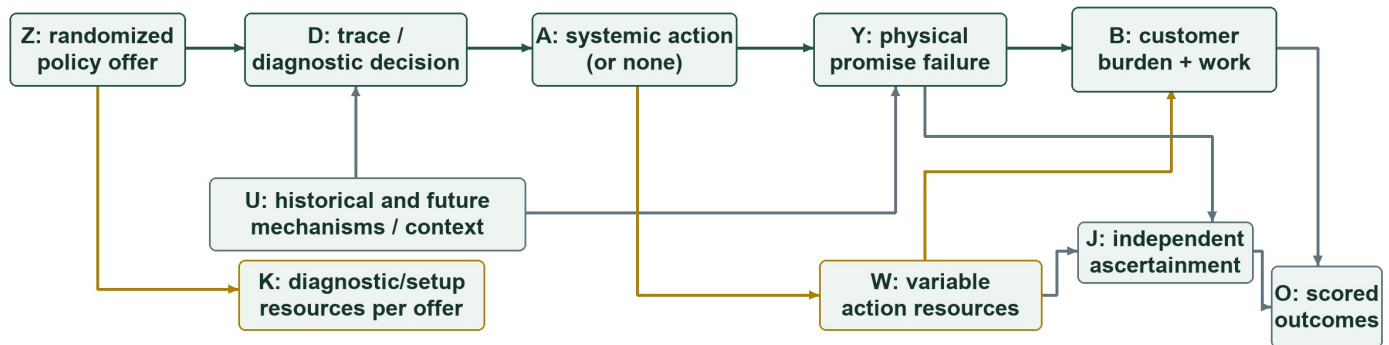
[AI/PD] The proposed Custody-to-Action Protocol (CAP) separates recorded actor identity, actual custody, a feasible systemic control and mature customer outcomes. A return label can guide a trace; it cannot, by itself, allocate causal responsibility.

OBSERVATIONAL LAYER: RETURN IDENTITY DOES NOT CLOSE THE CUSTODY PATH



C and Q affect selection; T and C can yield the same I. Conditioning on S changes the observed mix.

PROPOSED RANDOMIZED POLICY: DIAGNOSTIC COHORT TO FUTURE MANIFEST



K is charged once outside B. A no-new-action decision preserves current operations and customer remedies.

Reference J measures actual Y and W. Shared capacity and interference require trial boundaries.

[AI] In the historical layer, actual custody C and authorized transfers T can both affect recorded identity I. Capture quality Q also affects I and inclusion S in the observed return panel. Conditioning on S can change the composition of histories being compared. A compliant-sort label does not close the unobserved transfer path.

[PD] In the proposed trial, assignment Z changes diagnostic policy D and the selected systemic action A. Historical and future mechanisms U may differ; U affects both the diagnosis and future outcome Y. A changes Y and variable work W. Customer burden and W form B; fixed diagnostic/setup cost K is charged separately. Independent ascertainment J measures Y and W to produce O.

[PD] The randomized total effect retains all original future assignments, including domains selecting no new action. Do not adjust it for D, A, completed returns or successful links. Shared staff, routes and capacity require cluster boundaries; the graph is a testable causal proposal, not a recovered historical effect.

Basis: source role/selection distinctions [HR/AI]; proposed CAP policy trial [PD]. Arrows are hypotheses. Independent ascertainment means independent of the policy decision, not unrelated to actual outcomes.

TECHNICAL ANALYSIS

Identical return tables can imply different physical histories

[AI] An independent hypothetical table contains 100 returned shipments with confirmed marking and returning identities. The sort label is a recorded process classification. These examples are not transformed source data.

Observed class	Compliant sort	Other sort label
Same returning/markings actor	20	10
Different returning/markings actor	60	10

[AI] World A has no initial load-out errors. Every different-actor return follows an authorized rescue, reassignment or return handoff after a correct initial load. World B has initial allocation errors in the 60 different-actor/compliant cases; all other histories are unchanged. Both worlds reproduce every table cell, yet initial load-out error is zero versus 60%. Intermediate assignments are also compatible.

[AI] Thus a compliant-sort flag plus a different return actor does not identify initial custody or fault. The proof is stronger than a small-sample warning: even an arbitrarily large, perfectly measured table of these two variables cannot distinguish the two histories. Additional custody events or an intervention are required. The actual source is less resolved because its negative match category includes missing identity as well as a different identity.

[HR/AI] There is a second identity problem. One source compares the returning actor with the person who marked the stop; another compares the returning actor with the assigned person. In the hypothetical tuple assigned A, marking B, returning A, the first comparison is different and the second is same. Opposite labels can therefore be correct answers to different questions.

[PD] Preserve assigned, marking, loading, transfer, returning and scanning roles as distinct fields, with effective timestamps and authenticated handoff evidence. A person scanning at debrief need not be the earlier physical custodian. Trace approved rescue and reassignment before interpreting an identity mismatch as a prevention opportunity.

[AI] Reason motivates investigating conditions and defenses, but does not prove a station mechanism here. Fellegi and Sunter distinguish confirmed, rejected and unresolved record links; even a correct identity link does not establish the physical history between its endpoints.

Basis: independently enumerated observational-equivalence and identity-role proofs [AI]; Reason (2000), Fellegi and Sunter (1969). No historical error share or culpability is estimated.

TECHNICAL ANALYSIS

Name the eligible population before calculating coverage

[HR/AI] The narrow return panel, a broader defect taxonomy and a later replacement mart differ in role definitions, selection, period and key. Their denominators cannot be pooled simply because they describe the same operating topic.

Quantity	Required population and interpretation
Return identity share	Unique eligible returned shipment episodes with a specified pair of identity roles; separately show unknown and conflicting links.
Capture completeness	Reference-eligible returns with usable identity or reason evidence divided by all independently enumerated eligible returns.
Exposure-adjusted risk	Unique relevant defects divided by all originally assigned eligible shipments for that actor, route or domain; case mix and clocks matched.
Causal control effect	Difference in mature outcomes under defined alternative controls on a fixed original cohort; identity share is not this effect.

[AI] The narrow panel as a share of all defect rows is not automatically missingness among eligible returns: the complement contains no-return records. Exact row partitioning also does not prove disjoint entity partitioning. A shipment can contribute repeated events or appear across scopes. Validate the original shipment-episode key, duplicates, multi-actor links and the generating query before declaring package closure.

[AI] Suppose, hypothetically, a genuine eligible frame has verified coverage q and a validated binary share p among covered units. With no information outside coverage, the whole-frame share lies between $q \times p$ and $q \times p + 1 - q$. At $q = 0.25$ and $p = 0.80$, the range is 0.20–0.95. This bound is inapplicable if q uses the wrong eligibility denominator or p is merely an unvalidated causal label.

[PD] Sample missing, conflicting and unmatched records using known selection probabilities. Weight a validation sample only where each target stratum has positive observation probability and the linkage model is credible. Report unsupported strata separately. A high key-match rate to a high-value or defect-only table does not create a delivered-volume denominator.

[AI] Manski motivates explicit bounds when identifying assumptions are unavailable. Bareinboim and Pearl explain why population, regime and sampling differences matter when combining evidence. Neither source licenses transporting a selected return mix into a network-wide owner or benefit share.

Basis: source population forks [HR]; independently derived descriptive mixture bound [AI]; Manski (1990), Bareinboim and Pearl (2016). Historical coverage and causal responsibility remain unidentified.

TECHNICAL ANALYSIS

A scan trail and a customer promise are separate records

[PD] Build a versioned episode table that preserves original assignment, actual handoffs, observation times and the original customer promise. The process can recover a parcel while still missing its promise, or return it before the promise has expired.

Record layer	Minimum validation contract
Physical custody	Shipment episode; assigned, loaded, handed-off and returned roles; actual times; route/cart evidence and authorized transfer record.
Event capture	Scanning actor and device; event versus posting time; absent, late, corrected and conflicting evidence; raw taxonomy version.
Customer outcome	Original promise, confirmed receipt time, cancellation/nonreceipt and final follow-up status; preserve original cohort.
Accountability	Evidence custodian, next-action owner and governing scorecard rule stored separately from a causal-prevention assertion.

[HR/AI] The replacement analysis uses a different mart and window and reports a different reason mix. It is separately authored and cannot authenticate the earlier query outputs by agreement on a broad narrative. Access failure, source replacement and a recommendation change are evidence about the review process; they are not an independent replication of the original population.

[AI] Missing stow evidence admits at least physical non-containerization and recording failure. Concurrent movement across categories can arise from common processing changes, common operational shocks or both. A trend hold is defensible while these are tested; declaring the change an artifact from synchrony alone is not. Selected lifecycle examples illustrate possibilities rather than estimate their frequency.

[PD] Reconcile source-native definitions on a common linked cohort before comparing rates. Preserve zero, unknown, not applicable and unobserved states. Apply deduplication at the declared episode grain, not a blanket DISTINCT that deletes real repeat events. MAX over event flags may answer whether something was ever recorded; it cannot establish event order or exclusive states.

[AI] An eventual-delivery percentage at extraction is neither ultimate recovery nor on-time delivery. A fixed age recommendation needs validated event-age and revision curves. Censoring, replacement IDs and later return selection must remain visible. Customer remedies and financial accountability cannot be changed solely by a differently named operational bucket.

Basis: source role, lineage and time-window differences [HR/AI]; proposed temporal measurement contract [PD]. No original query, handoff history or customer-outcome panel is reconstructed from the register.

TECHNICAL ANALYSIS

Validate the custody trace and the action recommendation separately

[PD] A useful audit must distinguish a correctly linked actor, a supported physical history and a suitable future control. Agreement with the existing label tests none of those independently.

[PD] Freeze a codebook before sampling: confirmed correct handoff, confirmed misallocation, authorized later transfer, capture failure, other mechanism and unresolved. Define what each category requires and its observation horizon. Two reviewers independently inspect source evidence while blinded to the proposed owner and ranking. Adjudication can preserve unknown; it must not force every case into a blame category.

[PD] A random core of 200 independent episode traces gives an approximate 95% worst-case agreement screen of ± 6.9 percentage points. Clustered routes or shared capture devices reduce effective sample size. Supplement with difficult missing-identity, rescue and replan cases, then use sampling weights for population estimates. Kappa precision needs a prevalence-specific pilot; this screen is not a kappa sample-size calculation.

[PD] Report the full reference-by-label confusion matrix, raw agreement, kappa with uncertainty, unknown-truth share and category-specific false routing. Audit unmatched records as well as successful links. Measure correct action selection against independently supported operating evidence, not sponsor acceptance or a nominal owner field. Record actual intervention availability and the authority required to use it.

[AI] Wilson score intervals address sampling uncertainty for a defined binomial proportion. They do not fix selection into returns, multiple comparisons, shared dependence, absent exposure or incorrect identity roles. A lower-bound leaderboard is still a leaderboard of the measured selected fraction. Goldstein and Spiegelhalter support displaying adjustment and ranking uncertainty instead of implying that a precise rank identifies a preventable mechanism.

[PD] Validate the future-cohort mechanism separately from the historical diagnostic cohort. Capture route mix, transfers, handling changes and capacity before action selection. A policy with excellent historical trace accuracy can choose the wrong future control when the operating mechanism changes. No individual identity, punitive action or historical misconduct inference is part of this proposed research.

Basis: proposed blinded, probability-sampled validation [PD]; Cohen (1960), Wilson (1927), Goldstein and Spiegelhalter (1996), Fellegi and Sunter (1969). No historical reference accuracy or kappa is estimated.

TECHNICAL ANALYSIS

Diffuse counts do not establish a diffuse opportunity

[HR/AI] The source rejected individual naming after failing to find a relevant exposure denominator. That restraint is supported. Its stronger claim that singleton-heavy counts prove there is no useful coaching population does not follow.

[AI] Count concentration measures where observed events accumulated. Under a matched cohort and grain, a majority-singleton share of affected people coexists with a majority of events associated with repeat-event people. Multi-actor packages require reconciliation; neither statement identifies fault or coachability. Under a fixed-cohort, no-spillover scenario in which a control can prevent only events held by a selected group, a group with share s can remove at most s of those events. This is not a bound on its future risk or the benefit of a systemic control reaching all groups. High counts can reflect high exposure; low counts can coexist with high rates or a widely preventable mechanism.

[AI] Concurrent elevation across providers sharing a station is a clue to common conditions, not independent causal replication. Shared customers, route mix, recording, timing and capacity can create the pattern. Freeze a focus group before follow-up and compare mature outcomes with suitable untreated domains; do not call the most extreme observed group an established treatment opportunity.

[AI] Under an explicitly assumed linear conditional-mean model, expected untreated follow-up is $m + \rho \times (x - m)$, where m is the population mean and ρ is assumed persistence. Values 0.4, 0.7 and 0.9 imply 60%, 30% and 10% reversion of selected excess. These are quantitative sensitivities, not an estimated regression-to-the-mean fraction in this case.

[RC/AI] The frozen inventory contains 21 source/sensitivity screens: 18 weekly full/prefix/last-pair screens, two weaker dashboard/window screens and one attribution-age pair. Eight version audits and two design/goal audits are separate. All fitted X/MR points remain within tentative limits. A retrospective three-period-prefix holdout flags the final observation in four of six related metrics; it does not establish an operating cause. No stable phase is validated. Individuals limits use mean plus/minus 2.66 times mean moving range; the range upper limit is 3.267 times mean range. Conditional Laney calculations remain ineligible because binary identity, taxonomy, age and phase are unresolved.

[PD] An eligible future rate chart needs unique binary counts, real exposure, rational subgroups and stable measurement. Laney adjustment cannot repair changing taxonomy or selected return denominators. Prespecify moderators including transfer prevalence, identity capture, route type, future mechanism persistence and capacity. The full causal effect is not conditioned on postassignment returns, chosen action or diagnosis success.

Basis: recorded concentration and selection limits [HR/AI]; NIST/Laney process conventions, Barnett et al. selection/RTM. Exact source-screen inventory is retained privately and must be read with its eligibility qualifications.

TECHNICAL ANALYSIS

The contribution is analytical direction, not a proven operating gain

[HR/AI] Personal contribution is graded as a provisional registry-level hoop. The supplied register describes commissioning, analytical choices and an action proposal, but the named originals remain unrecovered in the bounded local search and operating execution is unestablished.

Claim and test	Supported boundary
Question and analysis direction: provisional hoop	The register links the focal leader to the ask, population choices, denominator challenge and action framing; source authentication is incomplete.
Query/document production: provisional hoop	Agent-assisted generation is described. Direction is distinguishable from personally executing every query or writing every line.
Replacement work: separate authorship	A teammate authored the later mart-based replacement. Recipient status and copied file metadata do not transfer authorship.
Adoption and outcome: unestablished	Owner-filled action rows, a proposed review cadence and a named test site do not show funded implementation or mature causal benefit.

[AI] The source also records an unapplied framing correction and an unverified scorecard composition. A recommendation can improve analytical discipline while still overclaiming causal ownership. Our present custody counterexample and reinterpretation are new case analysis, not accomplishments retrospectively attributed to the focal leader.

[AI] Consolidated identification limits are selection into the return frame, unobserved handoffs, no intervention comparator and no validated exposure-adjusted individual rates. Measurement limits are repeated entities, conflicting identity roles, missing capture, source-version breaks and incomplete promise linkage. A later source on a new population does not resolve those limits by narrative agreement.

[AI] Transport requires validated future mechanism and capacity assumptions. Economic value requires preventable mature outcomes, incremental work and non-overlapping costs; historical concession expense is neither a benefit estimate nor a complete service-burden measure. Public anonymization removes operational anchors, while private locators preserve reproducibility without claiming recovery of missing originals.

[PD] Close the historical record with the surviving decision: validate the custody mechanism and the population before choosing or reallocating a control. Unknown action adoption remains unknown beyond the supplied corpus; absence of a recovered completion record proves neither external success nor failure.

Basis: source chronology and contribution restrictions [HR/AI]; Collier process-tracing tests. Current analytical critique is distinct from the historical actor’s contribution.

TECHNICAL ANALYSIS

Test the complete diagnostic policy on a frozen future manifest

[PD] Randomize independent response-capacity domains to CAP or the approved usual diagnostic policy. Each policy may choose one of two authorized systemic packages or select no new action, which preserves usual operations. No customer or parcel is put on hold by this research choice.

[PD] The diagnostic cohort precedes action selection. CAP uses probability-sampled custody traces and an explicit uncertainty rule; usual practice uses its available return-label evidence. Freeze the subsequent original shipment manifest before assignment and before diagnosis is revealed, with equal planned exposure and eligibility in both arms. Every originally booked unit remains through transfer, diversion, nonreceipt and nonreturn. Preserve original promises, remedies and financial policies.

[PD] Primary ITT is usual-minus-CAP mean mature customer-failure cost plus variable action resources per original future shipment. Retain common reserved capacity in both arm totals; subtract extra fixed diagnostic/setup cost once in the investment comparison. The physical promise-failure ITT is a separate gate. Confirm receipt timing against the original promise by original promise + 35 days; unknown disposition receives sensitivity bounds on the original denominator.

[PD] The synthetic planning law lets one domain-level diagnosis choose a package for all future units. A historical mechanism and a later mechanism are jointly modeled; disagreement reduces action relevance despite a correct historical trace. Latent response types and classification accuracy are imposed planning assumptions: a custody trace need not identify counterfactual treatment response. Policy ITT can still be estimated from independent mature outcomes. Shared diagnosis makes shipment count alone a poor guide to precision. Independent-domain recruitment, transfer stability and reference quality need a pilot before the conditional numerical design becomes a field commitment.

[PD] Record diagnostic effort even if no action is selected, and action costs when attempts fail. Do not monetize scorecard transfer or fewer recorded returns as benefit. Independently monitor queue displacement, unsafe haste, ordinary-flow harm and customer remedy integrity; the worked outcome model does not automatically size these guardrails. Escalate adverse evidence under approved operating authority.

[PD] A later nonrandomized rollout could estimate group-time ATT against not-yet-treated domains, with no anticipation, common support and credible parallel trends. Use event-study and equivalence diagnostics, inspect case mix and spillovers, and power from a mature pilot panel. That future ATT is distinct from the randomized policy ITT and is not estimated here.

Basis: proposed CAP field evaluation and separately disclosed synthetic planning model [PD]; interference and Callaway–Sant’Anna methods for their stated designs. No historical intervention effect or operational feasibility is established.

TECHNICAL ANALYSIS

Four propositions make custody evidence accountable to outcomes

[AI/PD] CAP is a falsifiable decision framework. Better documentation, higher trace completion and a sharper label distribution are intermediate achievements; each must earn its connection to an appropriate control and mature customer benefit.

Proposition	Mechanism, boundary and disconfirming result
P1. Role-qualified linkage reduces false routing.	Separating assigned, marking, transfer and returning identities should reduce reference-audited incorrect owner hypotheses. It fails if added linkage merely increases confident labels while reference errors remain unchanged.
P2. Validated handoffs discriminate mechanisms.	Independent custody traces should separate authorized transfer, initial misallocation and capture failure. It fails if reference-audited errors do not decline as prespecified or unknown cases are forced into a cause. Distinct histories can still warrant the same control.
P3. Transport determines useful action selection.	Better historical diagnosis should improve future outcomes only when the mechanism and feasible control remain relevant. High trace accuracy with no mature benefit under changed future conditions falsifies automatic transport.
P4. Systemic value exceeds counting success.	A predeclared policy should improve the all-original-cohort outcome and clear its resource threshold while meeting guardrails. Lower return labels or more completed actions with unchanged physical outcomes fail this proposition.

[AI] The literature supplies complementary constraints. Record linkage preserves uncertain matches; human-factors and ranking research challenge actor-based blame and false precision. Deming warns against reacting to unstable variation. Campbell and Goodhart motivate testing whether targeting a return label changes recording rather than service. Ittner and Larcker require an evidenced chain from measured control to value; Tucker and Edmondson distinguish local workarounds from organizational learning. Composite-indicator guidance makes weights and compensability explicit: excellent reason coverage cannot compensate for unknown custody or failed safety. Partial identification exposes missing assumptions, and transport methods require a bridge to the future cohort. None supplies missing custody facts or validates this intervention.

[PD] Preregister the population, two action packages, uncertainty rule, reference, cost categories, inferential gates and safety margins. Preserve the original unknowns and source forks. Report independently reproduced planning results with their full assumptions; recalculate after pilot evidence changes capture, transfer, covariance or costs.

[AI] A failed mechanism test is useful when it prevents an unsupported control from becoming routine. The admissible decision is therefore an accountable experiment or further trace, not an automatic transfer of blame. Information value is modeled separately from operational benefit so the same prevented outcome is not counted twice.

Basis: new CAP propositions [AI/PD]; Fellegi–Sunter, Reason, Wilson, Goldstein–Spiegelhalter, Manski and Bareinboim–Pearl. These are testable proposals, not retrospective empirical findings.

A better historical diagnosis can still choose the wrong future action

[PD] Randomize independent station/provider capacity domains to the Custody-to-Action Protocol (CAP) or usual return-label interpretation. Compare policies for choosing approved systemic controls, with no individual sanctions or scorecard reassignment.

[PD] Freeze a probability-sampled historical trace batch and a separate future manifest of 240 originally booked units before assignment. Both arms have equal access and authority. An unresolved diagnosis preserves usual operations, not a parcel hold. Physical timestamps and receipt establish original-promise failures by original promise + 35 days; later receipt does not erase a missed promise.

Synthetic diagnostic law	Usual	CAP
Unresolved batch	5%	15%
Resolved accuracy / historical A	60%	85%
Resolved accuracy / historical B	30%	85%
Future mechanism persists / base	90%	90%

[PD] A/B are latent response types, equiprobable here; the future type persists or switches once per domain. Policies see traces, not the type. A trace need not identify counterfactual response in the field. Packages are handoff reconciliation or in-van search/dispatch discipline. Assumed future failure risks are 7% matched, 14% mismatched and 12% with no new package.

TRANSFER IS A TESTABLE CONDITION

[PD] If mechanism persistence falls to 50%, improved retrospective accuracy gives negative net value. Historical trace agreement cannot certify future package effectiveness. All original units, including diverted, nonreturned and unresolved units, remain.

[PD] Batch size and diagnostic calibration are unestablished; assumed accuracy is conditional on resolution, not derived from a trace count. All probabilities and costs are hypothetical.

Power the investment threshold and the physical outcome together

[PD] Exact finite mixture-binomial domain distributions preserve shared diagnosis and mechanism transfer. Independent domain sums calibrate strict one-sided gates at 2.5% each; a separate two-sided zero test has at most 5% error.

Design	Per arm	MDE (index)	BE detection	Joint floor
Initial	24	> full	8.2%	0.0%
Base candidate	576	5.04	82.1%	>99.9%
Transfer stress	576	5.04	82.2%	61.7%
Conditional redesign	1024	3.79	97.1%	85.9%

[PD] Break-even is 5.18 index points. Require 80% correctly directed detection at break-even and an 80% joint-power floor for a resource-benefit lower bound above incremental cost and fewer mature promise failures. The floor is $\max(0, P(\text{value}) + P(\text{service}) - 1)$, without endpoint independence.

[PD] The declared response path shifts domain probability from mismatched actions to matched actions and no new action. Each destination weakly lowers failure risk and work relative to mismatch; this coupling supports the composite-null boundaries. It is not an observed compliance measure or a claim about unknown field nuisance distributions.

[PD] Stress changes mechanism persistence from 90% to 75%. These are directly evaluated candidate allocations, not claimed global minima. More shipments within one misdiagnosed domain cannot substitute for independent domains.

CALIBRATION BEFORE RECRUITMENT

[PD] The conditional redesign requires 2,048 domains and 491,520 original future units. Recruitment is unverified. First calibrate trace validity, transfer, domain dependence and capacity effects; do not treat this numerical design as a deployment commitment.

[PD] MDE and costs use index points; base usual total modeled resource burden = 100. Delay, damage, queue spillover, admissibility and authority safeguards require separately sized gates.

Price the entire future cohort, including the cost of an unresolved trace

[PD] A promise failure costs 12 hypothetical resource units. Reserved capacity costs 0.10 per original unit in both arms, even without a new action. Executed package work adds 0.15; CAP adds $K = 0.08$ for every original unit.

[PD] K allocates trace sampling 0.035, independent linkage checks 0.025 and owner support 0.020. Package work is charged regardless of outcome; the common reservation remains in both arm totals. Research-only outcome adjudication has a separate study budget. No refund, scorecard charge or recorded reclassification is credited as a saving.

[PD] $Net_raw = a \times h \times 0.21384 + 0.015 - K$. Avoidability a and non-overlap h apply to favorable failure prevention only. Added harm and new work are charged in full. Table entries are net index points: $100 \times Net_raw / 1.54342$.

Avoidability a	$h = 0.50$	$h = 0.75$	$h = 1.00$
0.50	-0.75	0.98	2.72
0.75	0.98	3.58	6.18
1.00	2.72	6.18	9.64

[PD] At full realization, $K = 0.06 / 0.08 / 0.10$ produces 10.94 / 9.64 / 8.35 index points. Changed prices and transfer need fresh power. The model assumes work records are verified; if each arm has at most ϵ unknown or incorrect customer outcomes, loss-contrast uncertainty is 24ϵ and service-contrast uncertainty is 2ϵ .

A CORRECT TRACE IS NOT A BENEFIT

[PD] Retain missing outcomes and bound them; day 35 is a calibration assumption, not ultimate truth. Real field valuation needs randomized outcome differences, measured incremental resources and approved valuation over this same original manifest.

[PD] All prices, transfer probabilities and reference-error bounds are hypothetical. Concession spend is neither full service burden nor the avoidable value of the proposed controls.

PROPOSED INFORMATION VALUE / 4 OF 4

Buy evidence that can change which package deserves a test

[PD] A separate decision model partitions states by the preferred admissible validation: A, B or neither. Apply guardrails first and a frozen tie rule if both qualify. It does not assume that tracing or either package will improve operations.

Action / state	A supported	B supported	Neither
Validate A	13	-10	-6
Validate B	-8	15	-6
Retain / reconcile	2	2	5

[PD] Priors are 0.40 / 0.35 / 0.25. A signal reports the true state with probability q; each wrong state has probability (1 – q)/2. Payoffs include downstream validation expense. Evidence acquisition is additional, and unresolved mandatory evidence cannot authorize deployment.

Signal q	Gross EVSI	Acquisition	Net EVSI
0.50	1.0500	0.60	0.4500
0.70	4.2100	1.40	2.8100
0.90	7.3700	2.80	4.5700

[PD] Enumerating all 27 signal-contingent policies gives prior optimal value 2.7500 and EVPI 8.9500. Retaining operations is the prior optimum. Gross EVSI cannot exceed EVPI and is never added to the operating trial's modeled value.

DO NOT TURN A LABEL INTO A SANCTION

[PD] Evidence can redirect a process experiment or support retaining the current process. It cannot establish individual fault, authorize a provider penalty or replace the separate customer and operating safeguards.

[PD] Priors, signals and utilities are wholly hypothetical and separate from source counts, trial power and field value. No realized decision value is claimed.

Reproduce the population, the identity comparison and the decision

[AI] CAP separates three questions: what the record says, which mechanism could have generated it, and which feasible action improves the next cohort’s outcome. Agreement on the first question is insufficient to answer the other two.

Claim class	Meaning and replication boundary
[HR] Historical record	The register describes the initial assignment, analyses, action list and separately authored replacement. It preserves conflicting sources and unresolved ownership or measurement questions. It does not establish implementation.
[RC] Recalculated	Scope and grain reconciliations, cutoff-sensitive concentration, source-version arithmetic and tentative process limits. Constructive examples and prospective calculations are separately labeled; historical raw values remain private.
[AI] Analytical inference	A mixed return flag does not identify physical custody or cause. A high join match cannot validate the exposure denominator. Synchronized categories nominate a measurement investigation rather than proving a rule change.
[PD] Proposed design	The custody protocol, adjudication procedure, policy trial and economic scenarios. Their probability laws, costs and future service responses are hypothetical inputs requiring calibration.

Replication sequence

[PD] Recover and authenticate the original query outputs, code, correspondence and both document versions. Separate the initial owner’s analytical judgments from the teammate’s follow-up. Preserve machine-generated metadata as metadata; do not use it as sole authorship evidence.

[PD] Replay each filter at row and distinct-shipment grain. Verify overlaps, null semantics, identity-role comparisons, actual exposures, source windows and revision lag. Repeat calculations under every credible source version; retain unresolved forks rather than blending them.

[PD] Rebuild the alternative custody histories and the proposed trial from primitive inputs. Check diagnostic selection, future-mechanism transfer, original-manifest outcomes, all incremental costs and separate information value. Hash-bind the released source, numerical and actual-page reviews.

THE INTERPRETATION THIS CASE CHANGES

[AI] A return-identity mismatch is a lead for investigation. Turning it into a defensible intervention requires a custody trace, a verified action mechanism and a test on mature customer outcomes.

The separate private package contains source identities, raw values, exact locators, calculations and review records. The public case contains protected conclusions and disclosed methods.

RESEARCH REFERENCES / IDENTITY AND CAUSAL INTERPRETATION

Verify the link before interpreting the label

Primary research supports the interpretation and prospective design. It does not authenticate unrecovered project artifacts, identify historical fault or establish a financial return.

Bareinboim, E., & Pearl, J. (2016). Causal Inference and the Data-Fusion Problem. *PNAS*, 113(27), 7345–7352. DOI 10.1073/pnas.1510507113. Corrected author reprint. Populations, regimes and selection mechanisms determine what multiple datasets can identify. A larger secondary panel cannot validate a narrow return-identity panel merely because both concern the same operational question. Measurement identity must also be established separately.

Fellegi, I. P., & Sunter, A. B. (1969). A Theory for Record Linkage. *Journal of the American Statistical Association*, 64(328), 1183–1210. DOI 10.1080/01621459.1969.10501049. Primary publisher abstract and original article introduction. Linkage is an inferential classification problem with possible unresolved matches. A linked worker identifier is neither verified physical possession nor evidence that the worker caused the defect.

Goldstein, H., & Spiegelhalter, D. J. (1996). League Tables and Their Limitations: Statistical Issues in Comparisons of Institutional Performance. *Journal of the Royal Statistical Society A*, 159(3), 385–443. Author institutional manuscript, selected uncertainty and adjustment sections. Performance rankings require attention to uncertainty and comparability. A provider confidence bound is an investigation screen, not a causal verdict or a substitute for valid exposure.

Wilson, E. B. (1927). Probable Inference, the Law of Succession, and Statistical Inference. *Journal of the American Statistical Association*, 22(158), 209–212. DOI 10.1080/01621459.1927.10502953. Original article scan. Score inversion supplies the interval used in the historical ranking check. It addresses binomial uncertainty under its assumptions; it cannot repair unknown sampling, cluster dependence, invalid denominators or a mislabeled outcome.

Reason, J. (2000). Human Error: Models and Management. *BMJ*, 320, 768–770. DOI 10.1136/bmj.320.7237.768. Original article. A systems perspective motivates examining handoffs, conditions and defenses rather than assigning fault from an event label. It does not establish which loading, route or recovery mechanism generated this case's records, nor does it supply a logistics treatment effect.

Manski, C. F. (1990). Nonparametric Bounds on Treatment Effects. *American Economic Review*, 80(2), 319–323. Indexed original opening and bounds motivation. Bounds make maintained assumptions explicit. This case uses constructive histories to show why return labels do not identify physical fault, then distinguishes descriptive list coverage from a causal treatment response. No historical treatment-effect bound is estimated from an invented counterfactual sample.

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

Test the policy on the original future manifest

Primary research supports the interpretation and prospective design. It does not authenticate unrecovered project artifacts, identify historical fault or establish a financial return.

NIST/SEMATECH. e-Handbook of Statistical Methods, section 6.3.2.2: Individuals Control Charts. Official method. Compare tentative all-window and prior-window limits when a suspected category discontinuity changes dispersion. Four observations cannot establish a stable process phase. A crossing signals a question under the chart assumptions; it does not identify an operating or measurement cause.

Laney, D. B. (2002). Improved Control Charts for Attributes. *Quality Engineering*, 14(4), 531–537. Primary publisher abstract. Repeated counts and real denominators permit a provisional attribute calculation. Extra-binomial adjustment does not repair a changing taxonomy. Sparse or suspect-phase results remain mechanical screens rather than established control limits.

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20, 37–46. Primary publisher record. Independently code identity roles, custody transitions, authorized transfers and unresolved evidence. Report raw agreement, confusion tables, prevalence and uncertainty alongside kappa. Shared source errors can make two coders agree without validating physical truth.

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. Selecting the highest-count workers or most unusual provider/site pairs can generate apparent subsequent recovery. Freeze selection and compare repeated untreated trajectories; explicitly assumed reliability scenarios are not measured reversion.

Hudgens, M. G., & Halloran, M. E. (2008). Toward Causal Inference With Interference. *Journal of the American Statistical Association*, 103, 832–842. Original manuscript. Shared staging, dispatch, coaching and recovery resources couple operating units. The proposed trial groups those resources into independent domains; unmeasured spillovers remain a threat to the assumed assignment and variance model.

Callaway, B., & Sant'Anna, P. H. C. (2021). Difference-in-Differences with Multiple Time Periods. *Journal of Econometrics*, 225, 200–230. Author manuscript. A nonrandom staggered adoption requires explicit group-time effects, eligible comparisons and defensible parallel trends. It cannot be inferred from a post-adoption selected return share. This observational fallback needs its own power and sensitivity assessment.

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 an evidence purchase by the action its signal changes under a declared prior and payoff model. The separate information calculation is not historical savings and is not added to the policy trial's operating benefit.

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 the scorecard subordinate to the operating outcome

Primary research supports the interpretation and prospective design. It does not authenticate unrecovered project artifacts, identify historical fault or establish a financial return.

Collier, D. (2011). Understanding Process Tracing. *PS: Political Science & Politics*, 44, 823–830. Primary publication. Test the ownership of a specific analytical judgment, document version, adoption decision or implementation act. Machine-generated file metadata alone does not settle authorship. A derivative register warrants a provisional contribution grade until the original evidence is authenticated.

Tucker, A. L., & Edmondson, A. C. (2003). Why Hospitals Don't Learn from Failures. *California Management Review*, 45(2), 55–72. Author manuscript. Recurring workarounds and unowned system conditions can inhibit organizational learning. The case applies that mechanism to a trace-and-control design; it does not import health-care effect sizes or assume that systemic controls have already worked.

The W. Edwards Deming Institute. The Funnel Experiment. Official exposition. Adjustment to routine variation can worsen performance. A ranked exception list or short weekly movement requires a stable measurement basis before it becomes a recurring operating intervention.

Campbell, D. T. (1976). Assessing the Impact of Planned Social Change. Original institutional record. Using measures for accountability can change behavior and reporting. Track reason completeness, exclusion and appeals alongside physical service so a cleaner scorecard cannot substitute for a better customer outcome.

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. Return labels and compliance flags therefore remain distinct from custody, fault and preventable service failure.

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. Measures must connect to strategic choices and operating mechanisms. A complete action list, populated owner field or confident provider rank does not establish adoption or customer benefit.

OECD / European Commission JRC (2008). Handbook on Constructing Composite Indicators. Official handbook. Aggregation requires explicit definitions, weights and compensation assumptions. This case keeps economic clearance and the physical service gate distinct; improved diagnostic accuracy or lower administration cost cannot silently offset a failed customer 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 discipline; test the proposed response

[AI] This author-led, agent-assisted review is not external peer review or client approval. Its scope is a diagnostic operating case, new identification analysis and a proposed evaluation of an evidence-informed operating policy.

Major issue	Disposition in the revised case
Return identity treated as cause	Separated assigned, exception-marking and returning identities, including missing values. Constructed identical observed summaries with different custody histories; the loading explanation remains a testable hypothesis.
Panel robustness treated as validity	Kept duplicate sensitivity separate from measurement and transport. Preserved source versions, mixed null semantics, row/entity overlap and unresolved denominator differences.
Count concentration treated as treatment response	Limited concentration to a stated list’s observed coverage. Required exposure, recurrence and comparison evidence before individual attribution or claims about coaching effectiveness.
A synchronized jump treated as a known data change	Required version replay and independent event validation. Computed tentative process screens while retaining sparse-data and phase limitations. No statistical screen identifies the mechanism.
A better diagnosis treated as a mature benefit	Specified a prospective policy comparison on a separate original future manifest. Included uncertain mechanism transfer, full costs and customer-outcome testing; information value remains a separate model.

Partner critique: can a COO act on page one?

[AI] Yes: commission a bounded custody audit, reconcile the identity and exposure contracts, and require a testable control before expanding a coaching or station intervention. Assign the next investigation owner without declaring unverified fault. Keep individual penalties outside this diagnostic exercise.

Remaining conditions

[AI] Recover originals, reconcile source populations and scorecard authority, validate the independent custody reference, and calibrate the trial’s mechanism stability, costs, interference and feasible sample. Adoption and historical service or financial impact remain unestablished.

REVIEW DISPOSITION / PASS WITHIN STATED SCOPE

[AI] The diagnostic case and proposed CAP evaluation pass the stated internal review gates. This verdict does not establish historical causal responsibility, realized savings or readiness for unrestricted deployment.

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