

OPERATING CASE STUDY / CUSTOMER EXPERIENCE & LOGISTICS

Unit-Level Defect Diagnosis

A plausible routing mechanism is not yet an attributable customer failure

An operating-project case on a concentrated delivery-defect investigation. It separates route volume, planned stops, physical delivery and partner disputes, then defines the evidence and capacity required to test a remedy without changing the customer truth.

STOP GRAIN & ROOT-CAUSE ATTRIBUTION

BLUF / BOTTOM LINE UP FRONT

Verify the service unit and customer outcome before committing to a root cause.

RECALCULATED FINDING

Analytical finding: splitting an administrative stop can change a stop-weighted failure measure while every package outcome stays the same. Evaluate the original package population and the shared physical workload before calling the split an improvement.

CONTRIBUTION IN THE SOURCE RECORD

Provisional contribution: diagnostic authorship, commissioned review and specific evidence requests. Other teams owned operating changes.

EVIDENCE STATUS

Original artifacts remain unrecovered. The register describes draft analysis; dispatch, adoption, customer benefit and savings are unverified.

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

STRUCTURED ABSTRACT / DECISION BRIEF

Investigate the mechanism without inheriting its conclusion

[AI] The decision is whether a proposed change to delivery grouping addresses physical service failure, and which evidence can distinguish it from a change in reporting.

Context

[HR] A concentrated delivery-defect pattern reached an executive discussion about partner scorecard fairness. The recorded explanations included excessive grouping, forecasting, address data, field execution and an unavailable recovery path. The project owner became the named operational point of contact and commissioned a broader diagnostic with adversarial review.

Objective

[HR] Answer whether a large volume at one address had been consolidated into an unmanageable vehicle stop, identify executable countermeasures and retire unsupported assertions. The register reports draft deliverables and precise data requests. It does not establish that the reply was sent or that any change was implemented.

Approach

[AI] The Stop-Grain and Attribution Contract (SGAC) separates specification, effective configuration, realized grouping, physical execution, customer outcome and dispute adjudication. Reconcile the population at each transition. Preserve competing mechanisms, explicit decision rights and the distinction between a recorded analytic contribution and a delivered operating result.

Findings

[AI] Route totals need not equal a smaller population of approved disputed defects. Stop labels can change a metric without changing physical outcomes. A configuration log narrows the mechanism question but does not establish its causal contribution. High dispute approval describes selected adjudications; it neither proves upstream responsibility nor contains no information.

Decision value

[PD] Obtain a versioned stop-to-package crosswalk, original-promise outcomes and an authorized capacity test. Review scorecard fairness on its own evidence. Preserve true delivery status and customer remedies while investigating the operational cause. Separate the value of better evidence from any future operating benefit.

Implications

[AI] The owner's provisional registry-level contribution is diagnostic authorship, scope judgment, commissioned verification and specified requests. Stronger measurement can support a better decision before it supports a savings claim. Scale a tested mechanism only where the same constraints and service opportunities exist.

[HR] Selected source register S1. Original recovery, exact quantities, source versions and contribution evidence are in the private annex. [AI]/[PD] denote new analysis/design.

EXECUTIVE EXHIBIT 1 / COUNTING CONTRACT

A larger route total can contain a smaller defect population

[HR] The recorded diagnostic compared incident volumes with approved disputed defects and treated their inequality as a reconciliation failure. The populations had different grains and unresolved time links.

COUNTING CONTRACT / LINK UNITS BEFORE COMPARING TOTALS

Conceptual evidence chain: these quantities are related, but they need not be numerically equal



RECONCILIATION REQUIRES AN ENTITY CROSSWALK
More incident work than approved disputes can be entirely coherent; the missing link is identity and scope.

Source register S1 provides mixed-grain records; historical counts and exact versions remain private.

Quantity	What it can establish
Route or address volume	Work associated with an incident; it can include successfully delivered packages, other addresses and multiple stops. It is not automatically a defect count.
Original-package outcome	Whether a specific original booking met its promise, using a consistent reference and observation window. This is the customer outcome denominator.
Approved dispute event	An adjudication among submitted and eligible cases. Repeated submissions, reversals and reporting vintages may differ from unique packages.

[AI] The inequality alone proves neither corruption nor complete coverage. All, some or none of the disputed defects could belong to the incident packages until the identities and windows are linked. A missing crosswalk is an evidence gap; subtracting the totals does not measure unexplained losses.

EXECUTIVE ACTION

[PD] Reconcile original package identity, route, planned and executed stop, service date, promise, outcome and dispute lineage. Report unmatched records and duplicate joins rather than forcing unlike totals to agree.

[HR] Selected source register S1. Original recovery, exact quantities, source versions and contribution evidence are in the private annex. [AI]/[PD] denote new analysis/design.

EXECUTIVE EXHIBIT 2 / MECHANISM EVIDENCE

Configuration establishes a constraint, not its causal effect

[HR] The review distinguished system documentation from code semantics and withdrew some earlier claims. Historical runtime values and the realized grouping remained unavailable in the recorded analysis.

MECHANISM CONTRACT / DIFFERENT ARTIFACTS RESOLVE DIFFERENT CLAIMS

Proposed sequence: a valid configuration can establish feasibility without establishing causality



ONE LOG ENTRY CANNOT IDENTIFY THE ENTIRE FAILURE CHAIN
Trace the deployed constraint, the realized service unit and the original customer result separately.

Conceptual evidence structure; the arrows are not estimates of this project’s historical effects.

Evidence layer	Question it resolves
Effective build and configuration	Which rules, capacities and exceptions were available for this execution? A generic code path or test fixture is not a production value.
Realized execution	Which packages shared a vehicle stop and actual handling resources? Verify input, output, plan revisions, execution events and applicable service constraints.
Customer consequence	Did the executed grouping change physical completion by the original promise, compared with a feasible authorized alternative? A runtime log alone cannot answer this.

[AI] Several planned stops can still share an overloaded address or route. One large stop can also complete successfully. The one-stop-versus-many answer narrows the explanation; neither answer settles the full causal question. Cube, service-time, count and execution limits must be examined together.

EXECUTIVE ACTION

[PD] Give each evidence request a specific claim to resolve. Reconstruct the actual constraint and exposure first, then test the proposed physical response. Retain alternatives that the new artifact does not distinguish.

[HR] Selected source register S1. Original recovery, exact quantities, source versions and contribution evidence are in the private annex. [AI]/[PD] denote new analysis/design.

EXECUTIVE EXHIBIT 3 / DENOMINATOR INVARIANCE

A stop can split while the customer result stays fixed

[AI] A measure that gives every stop equal weight can change when a stop is administratively divided. The original-package failure measure survives a pure repartition.

PARTITION INVARIANCE / HOLD EVERY PACKAGE OUTCOME FIXED

Invented example: group A = 80 parcels / 40 failures; group B = 20 parcels / no failures

BEFORE: TWO STOP LABELS
Mean stop failure = $(50\% + 0\%) / 2$
25.00% stop average

AFTER: SPLIT B INTO 20 LABELS
Same group A; twenty successful labels
2.38% stop average

40 OF 100 ORIGINAL PARCELS FAIL IN BOTH REPRESENTATIONS
The label change improves the stop average while the parcel failure rate remains 40%.

Wholly hypothetical proof, not historical data or an index. Actual physical splitting needs its own service test.

Measurement choice	Decision implication
Average of stop failure rates	Each administrative stop has equal weight. Changing the partition can change the reported average even when every original package has the same result.
Failures / original packages	Counts each original booking once. Preserve its original promise and outcome reference; retain cancelled, reassigned and unresolved work.
Physical workload	Measure all handling, scanning, setup, travel and shared capacity. Smaller displayed groups do not automatically release physical time.

[RC] In the invented example, splitting only the successful group lowers the stop average from 25% to 2.38%; the same 40 of 100 original packages still fail. [AI] Real splitting could relieve a bottleneck or add harmful overhead. Those physical effects require a distinct treatment and resource test.

EXECUTIVE ACTION

[PD] Freeze the original-package denominator before assignment. Publish the stop measure as a diagnostic beside the package outcome; require evidence of physical benefit before crediting an improved stop average.

[AI]/[RC] Source-independent partition counterexample and algebra are disclosed in the technical layer and reproducibility files. No historical quantity is encoded in the exhibit.

EXECUTIVE EXHIBIT 4 / SELECTED ADJUDICATION

Approval is evidence about a selected process

[HR] The register reports unusually high dispute approval, incompatible headline versions and a dispute rule that appeared poorly matched to the failure state. It also records a request to improve the taxonomy.

ADJUDICATION SELECTION / PRESERVE THE PROCESS DENOMINATOR

Conceptual selection path: approval describes the reviewed set, not every underlying failure



FAIR PERFORMANCE REVIEW AND TRUE DELIVERY STATUS ARE SEPARATE
Investigate unsupported penalties while preserving customer rights and the physical delivery record.

Audit nonfilers, rejected cases, missing evidence and reversals; approval alone does not assign causal fault.

Tempting conclusion		What the evidence actually requires	
Upstream fault is established		Link the defect to its physical mechanism and evaluate responsibility against an effective rule. Approval alone does not identify who caused the outcome.	
Approval contains no information		Approval still describes the reviewed submissions. Its interpretation depends on filing, eligibility, selection, reviewer practice and available evidence.	
No-issue closure proves no problem		Inspect what the closure code means and which cases could be classified. A missing category can hide a recurring issue; the code does not independently prove that explanation.	

[AI] A fair correction to a partner scorecard and an accurate customer-delivery record are separate decisions. Removing an unsupported penalty does not authorize changing an undelivered package to delivered. A high individual defect count is also not sufficient evidence of individual fault.

EXECUTIVE ACTION

[PD] Provide an evidence-based review path without prescribing its verdict. Preserve status truth and customer remedies; audit both approved and rejected cases, nonfilers and unresolved records against an independent reference.

[HR] Selected source register S1. Original recovery, exact quantities, source versions and contribution evidence are in the private annex. [AI]/[PD] denote new analysis/design.

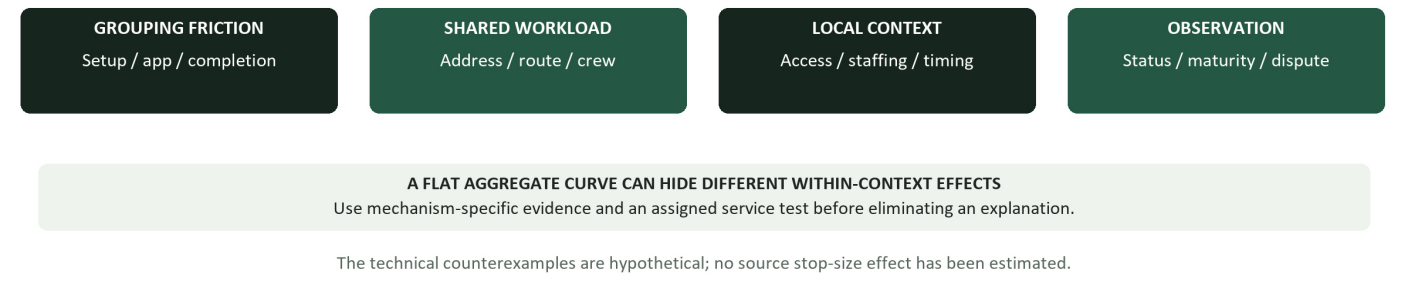
EXECUTIVE EXHIBIT 5 / DISCRIMINATING HYPOTHESES

A flat curve does not eliminate a harmful mechanism

[HR] The planned analysis would relate packages per planned stop to non-delivery. The recorded argument treated monotonic deterioration as support for consolidation and a flat curve as a reason to redirect the investigation.

RIVAL MECHANISMS / ASK WHAT WOULD DISTINGUISH THEM

Conceptual analysis: a crude stop-size curve combines exposure, context, selection and measurement



Observed pattern	Competing interpretation to test
Larger stops fail more often	Address difficulty, staffing, route load, timing or customer access may jointly determine size and failure. A descriptive gradient does not isolate the intervention.
A crude curve is flat	Different mixes of difficult and easy work can offset within-context deterioration. Limited range, caps, missing outcomes or measurement error can also mask an effect.
An incident follows an address fix	Verify deployment, applicability and the relevant feature at execution. Sequence alone neither disproves the address mechanism nor proves the fix failed.

[AI] Worst-site selection can exaggerate an apparent later recovery. Compare independent periods and suitable nonfocus units; assess whether the selected signal survives exposure, maturity and regression to the mean. Prespecify business line, node type and shared-capacity constraints as moderators.

EXECUTIVE ACTION

[PD] Use the observational curve to localize questions. Estimate the effect of a specific authorized action package with original outcomes, adequate independent domains and explicit ordinary-service protection.

[HR] Selected source register S1. Original recovery, exact quantities, source versions and contribution evidence are in the private annex. [AI]/[PD] denote new analysis/design.

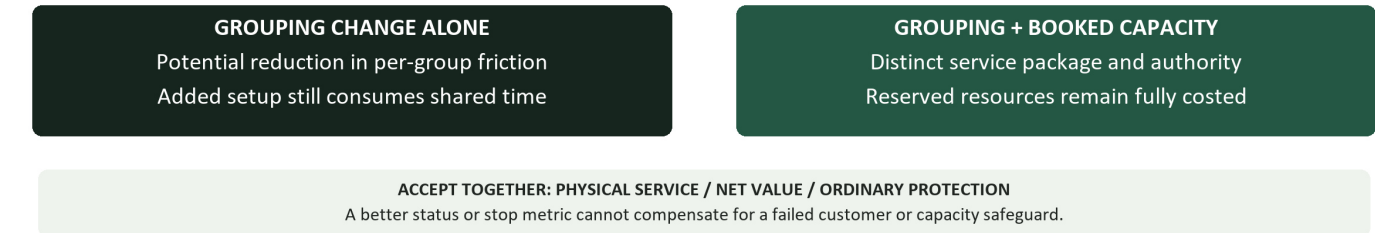
EXECUTIVE EXHIBIT 6 / REMEDY AND ECONOMICS

A smaller group needs a credible physical service budget

[PD] Compare current authorized practice with a defined change to grouping and capacity. Track physical delivery independently of status completion, while retaining the ordinary work sharing those resources.

SERVICE CAPACITY / EVALUATE WHAT THE PACKAGE ACTUALLY CHANGES

Proposed operating comparison: grouping, physical time, capacity cost and ordinary work are linked



Hypothetical evaluation architecture; a planning calculation is not evidence of deployment or realized benefit.

Operating test	Required acceptance evidence
Grouping change alone	Does reduced per-group friction exceed added setup and handling? Does ordinary service deteriorate when the same route absorbs the extra work?
Grouping plus dedicated capacity	Is the added resource available, authorized and fully costed, including idle, failed and late assignments? The capacity package is a distinct treatment.
Scaled deployment	Can the test detect economic break-even and meet customer and ordinary-service gates? Check reference validity, recruitment and local constraints before interpreting planning power as feasibility.

[AI] A larger trial cannot repair an adverse physical tradeoff. Count signed effects: target benefits, added failures, all incremental work and committed capacity. Cost each mature outcome once, using a finance-validated avoidable share and an overlap rule. A pallet or dedicated-route alternative still requires lineage, receipt, eligibility and customer-reference validation.

EXECUTIVE ACTION

[PD] Test the feasible resource package, not a label change. Keep expected information value separate from operating economics and do not book historical savings from an illustrative model.

[HR] Selected source register S1. Original recovery, exact quantities, source versions and contribution evidence are in the private annex. [AI]/[PD] denote new analysis/design.

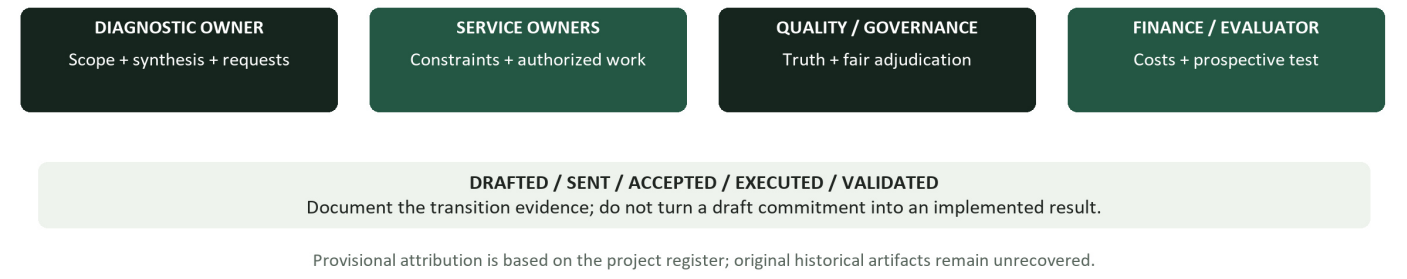
EXECUTIVE EXHIBIT 7 / CONTRIBUTION AND ACCOUNTABILITY

Own the diagnostic answer without claiming the cross-team fix

[HR] The register distinguishes the named point of contact’s authored analysis and critical-review decisions from the data, originating escalation, routing changes and dispute process owned by others.

CONTRIBUTION CONTRACT / KEEP OWNERSHIP ATTACHED TO THE WORK

Recorded diagnostic accountability and proposed cross-team closure are different evidence states



Party and decision right	Reviewable artifact
Diagnostic owner	Versioned brief; evidence ledger; retired claims; open alternatives; dated requests and acceptance status. Provisional registry-level attribution for authorship and commissioned verification.
Routing and operations	Effective constraint packet, stop-to-package lineage, approved service method, staffing commitment and physical execution evidence.
Quality and partner governance	Original-promise reference, taxonomy reliability sample, fair adjudication path and protected customer remedies.
Finance and evaluator	Bookable cost basis, overlap bridge, prospective test protocol and signed value assessment.

[AI] A source byline and a generated-file timestamp are different evidence. A draft request is not proof of dispatch, acceptance or execution. Critical reviewers can improve an argument without making every conclusion correct; the case rechecks their claims rather than inheriting their confidence.

CONTRIBUTION AND NEXT DECISION
[AI] The defensible contribution is a more precise diagnostic and accountable evidence path. [PD] Close the runtime, population and service-capacity questions before deciding which authorized remedy merits a bounded test.

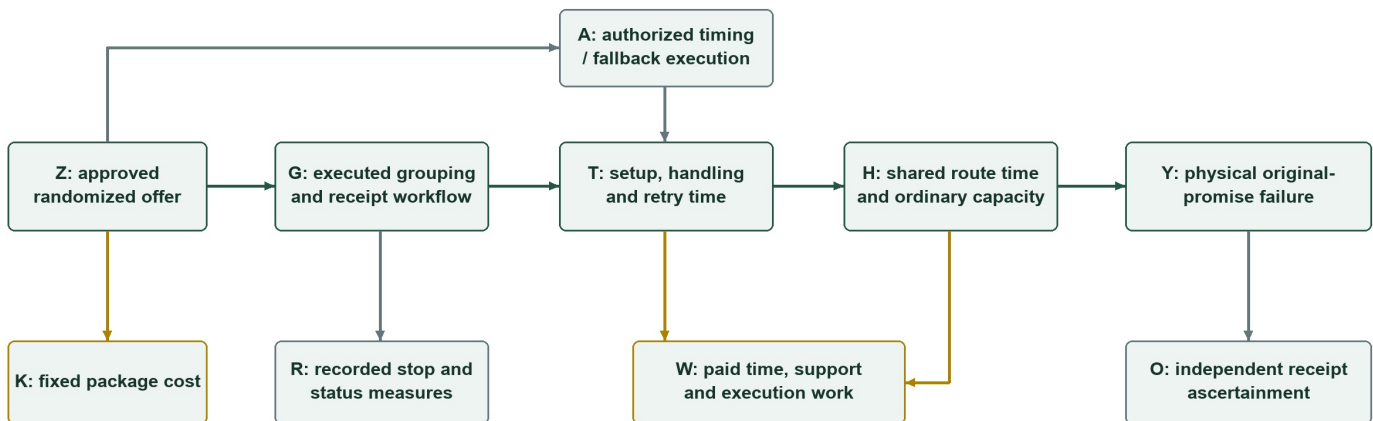
[HR] Selected source register S1. Original recovery, exact quantities, source versions and contribution evidence are in the private annex. [AI]/[PD] denote new analysis/design.

TECHNICAL ANALYSIS

Connect the executed service unit to the original promise

[AI/PD] The proposed Stop-Grain and Attribution Contract (SGAC) separates intended constraints, actual grouping, physical handling, recorded completion and independent customer outcomes. A configuration value or dispute decision cannot substitute for the full chain.

SGAC: GROUP LABELS, ACTUAL WORK AND CUSTOMER RECEIPT ARE DISTINCT



Reference quality J -> O

Baseline U affects G, A, T, H and Y. Randomization identifies the offer effect on original bookings.

G can change R even at fixed Y. R is a mechanism or administrative measure, not receipt itself.

Historical selection: Y -> S <- Q (filing / review propensity); other paths may also affect selection.

Z -> K. Shared capacity couples target and ordinary outcomes; keep both in the original denominator.

All arrows are proposed. Context, measurement and selection paths are stated separately for readability.

[PD] Assignment Z offers an authorized grouping and receipt workflow. It changes executed grouping G and setup/retry work T; physical handling still concerns the original parcels. Shared route time H can change both consignee and ordinary-flow original-promise failure Y. Independent reference O measures Y; recorded completion R is a separate downstream measurement.

[AI] Baseline U includes receiver access, package dimensions, route density, demand, service-time estimates, fleet configuration and reference availability. U can affect grouping and outcomes. Historical selection S depends on failure and filing, recording or adjudication Q; conditioning on approved disputes can distort associations among possible causes.

[PD] Randomize independent shared-capacity domains before execution, retain every original booking and keep authorized remedies unchanged. Actual group count, retries and completed scans are mediators, not adjustment variables for the total offer effect. Logs establish recorded execution only within validated coverage. Arrows represent proposed mechanisms, not established historical causes.

Basis: SGAC theory of change [AI/PD]; source inference boundaries [HR]. Cole et al. (2010), Kahan et al. (2023), Sigelman et al. (2010).

TECHNICAL ANALYSIS

Reconcile opportunities, failures and selected disputes

[HR/AI] The record compares address-level or route-level package volumes with a selected approved-dispute population. A larger opportunity count is compatible with a smaller outcome subset; the inequality alone establishes neither a contradiction nor an attribution.

Object	Required grain and linkage
Booking / parcel	Original unique parcel, customer deadline and route assignment before the offer.
Planned / executed group	Plan revision, service-group membership, execution order, receiver and actual grouping version.
Physical outcome	Independently supported deadline receipt, later receipt or unresolved state, with extraction vintage.
Administrative outcome	Lost label, disputed defect, approval and financial line item kept as distinct event histories.

[HR/AI] The historical opportunity figures also mix route, address and individual-delivery wording. Even their sum is not a verified unique-parcel denominator. Only an entity-level join can show how many approved defects fall within those opportunities. Larger totals do not prove that the cited dates fail to contain all selected defects.

[HR/AI] A reported approval fraction describes decisions among filed cases under a particular rule and evidence standard. Conflicting source versions remain unresolved. High approval neither estimates underlying defect prevalence nor proves upstream control. An apparently incompatible reason-code requirement raises a rule/application question; it does not make every reviewer decision uninformative.

[PD] Freeze source, effective rule, entity keys, plan-created and execution dates, as-of vintage, filing eligibility and duplicate handling. Test one-to-many joins explicitly. Preserve legitimate line items for valuation instead of applying a blanket distinct-count rule. An empty partition is missing evidence until its availability and coverage are verified.

[AI] A transporter-linked outlier, a closed ticket or a dispute theme is not an identified physical cause. No individual fault or complete network mechanism is established by this selected record.

Basis: source grain, selected approval and version conflicts [HR/AI]; proposed reconciliation contract [PD]. No historical quantities or identities are disclosed.

TECHNICAL ANALYSIS

A better stop metric can leave every parcel unchanged

[AI] Repartitioning the same parcels can change an unweighted average across stops without changing physical service. The example below is independent of the historical data and changes labels only.

Synthetic partition: 100 fixed parcels	Mean stop failure fraction	Parcel failure
Two stops: 80 parcels / 40 failures; 20 / 0	25.00%	40.00%
Split only the successful stop into 20 one-parcel labels	2.38%	40.00%
Instead split the 80-parcel stop into single-parcel labels	49.38%	40.00%

[AI] For a partition into S nonempty stops, the unweighted stop mean is $(1/S) \cdot \sum_s (d_s/n_s)$, where n_s is parcels and d_s failures. The original two-stop mean is $(0.5+0)/2$. After splitting the successful stop, it is $0.5/21=1/42$. All 40 failed parcels remain failed.

[AI] In contrast, $\sum_s n_s \cdot (d_s/n_s) / \sum_s n_s = \sum_i Y_i / N$, where N is original parcel count and Y_i the parcel failure indicator. With an exhaustive disjoint partition and unchanged outcomes, the parcel-weighted risk is invariant. The proof checks every feasible number of successful-stop fragments and a partition that worsens the stop metric.

[PD] Prespecify whether the target is an average parcel, an original receiver opportunity or an administrative stop. None is automatically the other. A grouping intervention changes the number and composition of stops, so post-treatment stop counts cannot replace the fixed original manifest in the customer estimand.

[AI] Kahan and colleagues motivate matching weighting to the target under informative cluster size. SGAC adds a case-specific operational concern: the policy can create the reporting groups themselves. Track grouping as a mechanism measure while funding decisions use independently ascertained original-parcel outcomes and total resources.

Basis: exact synthetic repartition proof [AI], counterexample.py/json; Kahan et al. (2023). These rates are hypothetical and have no mapping to source values.

TECHNICAL ANALYSIS

Neither a flat curve nor a runtime ceiling settles cause

[AI] Stop size can covary with receiver access, route demand and planned service time. A crude size-risk curve is an association. A flat curve need not refute harm; a rising curve need not prove that changing grouping will help.

Independent structural example	Easy context	Hard context	Overall
Small grouping	50 parcels; 10% fail	150 parcels; 30% fail	50/200 = 25%
Large grouping	150 parcels; 20% fail	50 parcels; 40% fail	50/200 = 25%

[AI] The declared construction gives each parcel two potential outcomes using the same balanced latent decile. Larger grouping adds a failure for one extra decile in either context: a 10-point causal increase within each context and overall. Different context mixes cancel it in the crude curve. The causal conclusion follows from the constructed potential outcomes, not from stratification alone.

[AI] A parameter log is similarly limited. In a second synthetic construction, 20 parcels pass the same count ceiling of 100 and estimated-service ceiling of 10, with estimated service 4. Actual service can be 4 or 16 against an unchanged deadline of 12. The logged values and planned feasibility agree while customer outcomes differ.

[HR/AI] Several stops at one address can still share a bottleneck, contain an oversized group or incur repeated setup. Conversely, one large group is not sufficient proof of harm. Missing a universal count cap does not remove cube or service constraints. A source-code interpretation also needs the historically deployed build, configuration and inputs.

[PD] Request linked plans, actual memberships, service estimates, realized work, receipt and versioned logs. Test a bounded approved response on original bookings. Competing forecasting, access, handling and recording explanations can coexist; one narrow artifact cannot select a complete causal story.

Basis: two new exact constructions [AI]; Simpson (1951), Solomon (1987, abstract scope), runtime/source distinction [HR/AI]. All displayed numbers are synthetic.

TECHNICAL ANALYSIS

Validate the trace without treating it as physical truth

[PD] A useful trace must connect booking, planning, execution and independent receipt at the right grain and time. More logged fields do not automatically improve linkage validity, coverage or the customer outcome.

Measurement layer	Validation requirement
Runtime exposure	Build/configuration hash, constraint units, plan version, applied strategy, complete member IDs and execution clock.
Actual work	Physical handling, retries, setup, authorized support, denied or unavailable options and paid fallback.
Receipt and status	Original-deadline physical receipt separately from scan upload, Lost posting, disputes and financial reversals.
Coverage and version	Known inclusion probabilities, missing partitions, duplicates, cross-day links, clock alignment and backfill.

[PD] Draw a stratified probability sample from original bookings, including unfiled, denied and approved cases, different group sizes, receiver types and baseline route conditions. Two coders blinded to assignment and later outcome classify membership and action applicability. A separate adjudicator uses time-stamped physical reference evidence; retain unresolved cases.

[PD] Report population-weighted confusion matrices, prevalence, raw agreement and kappa with domain-bootstrap intervals. A simple independent sample of 400 gives worst-case approximate 95% precision of plus or minus 4.9 percentage points for one agreement proportion; it is not kappa power. Rare wrong links, clustered observations and oversampling require their own precision calculation.

[AI/PD] NIST logging guidance motivates consistent identifiers, timestamps, retention and protected versions. Dapper motivates connected execution traces and explicit sampling. Neither establishes physical custody or receipt. Test error invariance across versions and baseline contexts, and blind outcome adjudicators to the operational hypothesis.

[PD] Preserve original denominators. Unknown receipt is neither automatic success nor automatic failure: report worst-case bounds and a sensitivity model. An administrative correction cannot retroactively create a delivery or erase a missed original promise.

Basis: proposed SGAC measurement contract [PD]; Kent and Souppaya (2006), Sigelman et al. (2010), Cohen (1960). No historical logging implementation is claimed.

TECHNICAL ANALYSIS

Separate recurrence, selection and transportability

[HR/AI] Selected incidents can establish that a concern recurred, subject to source authentication. They do not establish a stable process rate, a structural causal mechanism or a reliable ranking of controllable loss.

[RC] The final audit contains 3 source/sensitivity screens, 9 separate version audits, 10 design-shape audits and 8 profiles. The 3 source screens stay within tentative individual and moving-range bands, but mixed grains and unequal gaps prevent process interpretation. Two tier-design profiles have arithmetic crossings, not operating signals. There is no validated stable phase and no eligible or computed Laney panel; repeated binary counts and denominators are unavailable.

[PD] For genuinely comparable individuals, $MR_t = |x_t - x_{(t-1)}|$. Tentative individual limits are $\text{mean}(x)$ plus or minus $2.66 * \text{mean}(MR)$, with moving-range upper $3.267 * \text{mean}(MR)$. Sparse or mixed-grain screens remain non-diagnostic; within-limit points do not establish stability. Laney requires valid repeated binary counts and exposures, not reconstructed counts from rounded approval rates.

[AI] Worst-site selection also invites regression to the mean. Under an assumed linear conditional-mean model, expected untreated follow-up is $m + \rho * (x - m)$: m is population mean and ρ persistence. Assumed ρ of 0.4, 0.7 or 0.9 implies 60%, 30% or 10% reversion of the selected excess. These are hypothetical sensitivities, not estimated historical fractions.

[PD] Prespecify baseline receiver availability, original volume, package dimensions, vehicle capacity, route density and authorized fallback readiness as moderators. Actual group count, retries and receipt completion are postassignment mediators. A small subgroup or selected high-approval cohort needs uncertainty, not a causal league table.

[AI/PD] A documented geographic edit preceding an incident does not establish effective deployment or exclude remaining access constraints. Test the same approved response across prospectively defined contexts. A dedicated channel may remove one bottleneck while creating another; its staffing, eligibility, service and ordinary-flow effects require separate evaluation.

Basis: source recurrence and selection limits [HR/AI]; proposed variation/transport protocol [PD]. NIST, Laney, Barnett et al.; historical coordinates remain private.

TECHNICAL ANALYSIS

Separate analytical ownership from operational authority

[HR/AI] Contribution is a provisional registry-level hoop: a named contact, scope escalation, commissioned analysis, corrections to earlier claims and owner-addressed evidence requests. No sent reply, accepted remediation or mature impact is authenticated by this record.

Claim object	Supported boundary
POC and analytical direction	Registry records responsibility for the answer and machine-assisted research; originals still require authentication.
Draft artifacts and claim retirements	Supports analytical production and stated judgment; file existence does not establish delivery or influence.
Routing, support and adjudication changes	Owned by other accountable roles; proposed asks do not constitute approval or implementation.
Financial and customer results	Unmeasured; retired prices, staffing estimates or refund assertions cannot become realized benefit.

[AI] Multiple agent passes can challenge claims but do not create independent source evidence. Code takes precedence over a wiki only for the code and version actually inspected; it does not authenticate historical runtime. A failed data request is an evidence limitation, not proof of absent events or an exoneration.

[HR/AI] Only specifically documented closure codes can be described as observed. A no-issue label is not a complete causal investigation, while missing a named cause code does not erase every signal. The proposed taxonomy must allow unresolved evidence rather than presupposing consolidation as the cause.

[PD] Treat fair dispute review, recorded status, physical delivery and customer remedy as separate decisions. A scorecard exemption need not itself alter receipt or entitlement. A false delivered update can harm customers; an authorized correction supported by receipt evidence is a different action. Preserve the applicable customer remedy while evidence is unresolved.

[PD] Ask each owner for a versioned artifact, decision threshold, accepted fallback, due date and execution acknowledgment. Measure whether the analytical packet changes a defensible decision before claiming realized decision value. A longer document or larger review fleet is not an outcome.

Basis: recorded contribution and authority [HR/AI], Collier process-tracing logic; proposed decision rights [PD]. Partner and customer interests require separate evidence.

TECHNICAL ANALYSIS

Test the service process on a fixed original manifest

[PD] Compare an authorized grouping and receipt package with current approved practice at independent route-capacity domains. The intervention may change setup, retry and handling time; it cannot improve its customer denominator by creating more stop labels.

Element	Prospective specification
Population	Externally scheduled original parcels and promises, including consignee and ordinary-flow units and any dedicated-path transfer.
Treatment	Specified approved grouping/receipt workflow; unchanged remedy rights; no unsupported delivered-status shortcut.
Dependence	Repeated route-days share domain context and capacity; parcels and days are not independent randomized clusters.
Decision	Original-parcel physical improvement, resource gain above fixed cost, and a prespecified ordinary-flow harm margin.

[PD] Let $Y_i(z)$ be original-promise failure under offered version z . Estimate $E[Y_i(0)-Y_i(1)]$ on all original parcels; estimate net resource benefit on the same population. Independently reconstruct deadline receipt by the proposed original-promise-plus-21-day reference. Validate that horizon and retain unresolved outcomes; eventual receipt does not erase an earlier miss.

[PD] The new finite time-budget model retains physical handling under repartition and charges setup/retry work. Exact domain moments and conservative variance bounds support explicitly approximate normal planning. Direct finite-process simulations use actual physical-null, resource-zero, economic and safety-boundary laws, with Monte Carlo intervals. These hypothetical known-law bounds require field calibration; no all-nuisance finite coverage or globally minimal recruitment is claimed.

[PD] Compare an 80%-power MDE with economic break-even in common units and horizon. Evaluate joint gates and stress capacity assumptions. More observations cannot rescue a harmful mean; an earlier staffed appointment or dedicated handoff is a new costed version. Discount only favorable avoidable, non-overlapping benefit; charge added work and ordinary harm in full.

[PD] A later observational rollout needs a separately defined group-time ATT, credible controls, spillover and positivity checks, pretrend plots plus equivalence/sensitivity analyses. Nonsignificant pretrends do not prove identification. EVSI remains a separate evidence-choice model, not additive operating savings.

Basis: proposed SGAC randomized original-manifest ITT and finite planning model [PD]; Kahan et al., Solomon, interference methods and Callaway-Sant'Anna. All parameters are hypothetical.

TECHNICAL ANALYSIS

SGAC makes the next four decisions falsifiable

[AI/PD] The framework tests linked exposure, mechanism, customer service and decision value separately. Each proposition can fail without turning uncertainty about the historical root cause into a claim of no operational problem.

Proposition	Prespecified disconfirmation
P1: Runtime and outcome grains reconcile	Independent entity/version audits fail membership, coverage or clock agreement targets.
P2: The package improves service work	Randomized offer fails the intended reduction in avoidable setup/retry burden or misses the before-promise completion target.
P3: Improvement reaches original customers	Original-parcel receipt fails to improve, or ordinary-flow harm exceeds its approved margin, despite better stop or dispute metrics.
P4: Value survives operational context	Full-resource benefit fails break-even, or prespecified receiver/capacity strata fail out-of-sample transport criteria.

[AI] The repartition proof shows how administrative grouping can change a stop statistic without physical progress. The flat-curve construction defeats a proposed universal falsification rule; the identical-log construction defeats parameter-to-outcome sufficiency. These are new analytical contributions, not reconstructions of the historical incidents.

[AI] Kahan connects the estimand to its weights; Simpson and Cole distinguish aggregation and selection. Solomon motivates joint scheduling constraints. Dapper and NIST motivate trace linkage and coverage, while SQUIRE separates intervention description from studying whether it works. None supplies this project's missing causal effect.

[AI] Deming cautions against tampering after selected spikes. Campbell and Goodhart predict adaptation of targeted labels; Ittner-Larcker-Randall require a tested action-to-outcome chain. Tucker-Edmondson motivates examining workarounds as responses to process constraints. Composite resource weights must not conceal ordinary-flow harm or unfair adjudication.

Limitations and identifying assumptions

[AI/PD] Original policy and runtime versions remain unrecovered. No joined original-parcel panel or credible historical counterfactual is available; selected disputes and negative searches cannot identify the cause. The prospective service, receipt-error and calendar laws are hypothetical and need calibration. Recruitment, receiving-window feasibility, interference and authorized treatment availability can invalidate the proposed design. Resource prices and production confirmation costs need validation; research-only adjudication is separate. Archive versioned claims, exact proofs and final rendered hashes, but do not equate reproducibility with identification. This case supports reconciliation and a bounded approved test, not historical implementation, customer impact or realized cash.

Basis: proposed SGAC propositions and synthesis [AI/PD]; six fresh primary sources, SQUIRE and continuing methods. Reproducibility does not substitute for identification.

Splitting does not create handling capacity

[PD] Randomize independent shared-crew and consignee-capacity domains to current approved grouping or a specified grouping-and-receipt package. All inputs below are hypothetical; no historical effect or system limit calibrates them.

[PD] Five externally scheduled days per domain each retain 60 target-address and 120 ordinary original parcels. The offer replaces one group with three groups of 20. Every parcel still needs one service minute. Sequential setup and retries consume time before handling; unfinished parcels remain original-promise misses.

Planning input	Explicit finite law
Persistent domain grade	Equal chance of -5 or +5 minutes across all five days
Target / whole-route time	80 / 198 minutes, each shifted by the same grade
Setup per attempted group	6 minutes base; 10 minutes in friction stress
Retry: usual / grouped	50% chance of 30 minutes / 8% chance of 10 per group
Receipt after timely service	98% independently for each served parcel
Staffed redesign	30 paid earlier minutes; original promises unchanged

[PD] After target work, remaining route time serves ordinary parcels. Unserved, diverted and late-recovered parcels stay in the denominator. Independent reference through original promise + 21 days reconstructs physical receipt; application completion and dispute approval are not receipt truth.

[PD] BUY A DIFFERENT OPERATING CONSTRAINT

The staffed package is available only inside a preexisting approved receiving window. All reserved crew minutes are paid even unused. This is a changed treatment, not a larger sample or a retrospective promise extension.

[PD] Constructed timing and receipt law. Stop counts never replace the 180 original parcels per day. Perfect eligibility/linkage is a calibration assumption.

PROPOSED OPERATING EVALUATION / 2 OF 4

Size the test, then fix the failing mechanism

[PD] Require positive economic lower bound above cost, original-parcel physical improvement, and ordinary-flow noninferiority. The candidate +0.6 percentage-point harm margin among ordinary parcels requires Quality approval; it is not established safety.

[PD] Exact finite time/retry moments preserve five-day shared-grade dependence. Normal critical values use variance bounds over a declared domain-response mixture; directional bounds are nominal 97.5%. The zero test is two-sided 5%, but MDE and break-even detection count only beneficial-direction rejection.

Direct design	Per arm	MDE index	Cost index	BE detection	Joint normal floor
Initial	16	-	12.55	16.2%	0.0%
Base candidate	160	11.36	12.55	87.3%	99.7%
Setup-friction stress	160	-	12.55	-	0.0%
Staffed redesign	128	16.08	21.19	95.8%	>99.9%

[PD] Initial allocation cannot detect break-even. The base candidate has MDE below cost and passes the three-gate planning rule. Stress makes expected net value negative and ordinary harm +0.817 pp, above the candidate margin. No sample size repairs that mean; the staffed version changes the time budget. A zero joint floor is uninformative, not zero joint probability.

[PD] Eight independently seeded finite-process runs use 6,000 trials each. Full joint success: base 99.57% (Wilson 95%: 99.37-99.70%); staffed 6,000/6,000 (99.94-100%). Economic-boundary false clearance is 2.47% / 2.28%; ordinary-boundary false clearance 2.58% / 2.32%. These are simulation estimates, not uniform coverage guarantees.

[PD] CALIBRATION PRECEDES RECRUITMENT

Base testing would need 320 independent domains and 288,000 original parcels; feasibility is unestablished. Measure receiving windows, retry law, reference error and covariance first. The response mixture is a planning path, not observed adoption.

[PD] Same fixed base-usual resource reference = 100. MDE and costs are index points. Joint floor = max(0, sum of three normal-planning powers - 2).

Charge the route, including unfinished work

[PD] Base usual burden is 0.597667 hypothetical resource units per original parcel, indexed to 100. Each promise miss carries an assumed 8-unit burden, including later handling and remedy costs; this is not historical cash.

[PD] $Net_raw = a * h * \text{favorable physical-harm reduction} - \text{added physical harm} - \text{added variable work} - K$. Avoidability a and benefit non-overlap h discount only favorable components. Added harm and all work stay fully priced. $Net_index = 100 * Net_raw / 0.597667$.

[PD] Common crew capacity costs 0.04 per reserved minute, including idle time; each attempted group costs another 0.30 in disjoint support resources. Base $K = 0.075$ per original parcel: 0.025 each for integration, production receipt control and coordination. Staffed $K = 0.126667$ adds 0.045 appointment management plus 30 paid minutes per day. Research-only adjudication is a separate budget.

a / h	K factor	Base net	Stress net	Staffed net
1.00 / 1.00	1.00	+21.03	-32.94	+57.92
0.75 / 1.00	1.00	+12.09	-32.94	+38.00
0.50 / 0.80	1.00	-0.42	-32.94	+10.12
1.00 / 1.00	1.25	+17.89	-36.07	+52.62

[PD] Base components per original parcel are 0.21365 target benefit, 0.00964 ordinary harm, 0.00333 extra support and 0.075 fixed cost. A pure stop relabeling receives zero physical benefit. Reprice the full signed outcome vector for a real trial, including fallback, late handling and overtime.

[PD] A LATER RECORD IS NOT A LATER SUCCESS

Validate day-21 maturity without moving the promise. With 0.5% uncertain outcomes in each whole-manifest arm, physical contrast error can reach 1 pp and burden error 0.08 units. Ordinary-harm bounds require uncertainty within each original ordinary stratum; unknown stratum membership needs separate bounds.

[PD] Net index points; same original-parcel cost pool. Outcome-error bounds assume verified resource costs. No annualization or realized savings.

Learn which constraint deserves a trial

[PD] Evidence can select a grouping test, a staffed-appointment test or further validation. This separate utility model remains meaningful when operating benefit is unproven and cannot be added to physical trial value.

Hypothetical utility payoff	Group transaction friction	Physical receiving capacity	Unreliable outcome evidence
Retain and validate	2	1	7
Test grouping package	9	-6	-5
Test staffed appointment	5	10	-7

[PD] State priors are 40% / 35% / 25%. Prior action utilities are 2.90 / 0.25 / 3.75; EVPI is 5.10. Each signal names the true state at the stated accuracy and divides errors equally across the other two. All 27 signal-contingent action policies are evaluated.

Signal accuracy	EVSI	Evidence cost	Net utility
40%	0.0950	0.3500	-0.2550
70%	2.1525	1.2000	0.9525
90%	4.1175	2.4000	1.7175

[PD] The weakest signal costs more than it is worth. Stronger signals choose the grouping test after a transaction-friction signal, the staffed test after a capacity signal, and further validation after an unreliable-evidence signal. Assumed decision payoffs are not operational savings or a mandate to deploy.

[PD] AUTHORIZE THE EVIDENCE, THEN THE VERSION

First verify actual constraints, independent physical receipt and approved appointment access. A later randomized test estimates original-manifest policy ITT; no individual component effect follows from a bundle comparison. Known safety defects must be corrected before a live trial.

[PD] Separate three-state Bayesian information model. Evidence acquisition is subtracted once. EVSI cannot exceed EVPI; neither is added to the trial economics.

REPRODUCIBILITY / READER PROTOCOL

Reproduce the unit, then test the service response

[AI] SGAC connects the original package population to executed service units, shared physical capacity and independently measured customer outcomes.

Reproduction layer	Required reconstruction
Source and attribution	Match selected cells to the immutable register. Preserve recovery attempts, versions and ownership boundaries. Authenticate original bylines and communication separately if recovered; do not infer dispatch from a draft.
Population and grain	Link route and address work to unique original packages, effective stops, service dates, outcomes and dispute events. Preserve unmatched records and duplicate joins. Unequal totals alone do not identify an error.
Mechanism and measurement	Reproduce the pure-repartition identity, mixed-context counterexample and runtime-log boundary. Verify which constraints operated, where work shared capacity and what the customer actually received.
Evaluation and economics	Rebuild the finite service law, direct trial allocations, effect thresholds and selected simulation checks. Retain original-promise failures, ordinary work, every incremental resource and signed harm.
Information and decisions	Recompute all contingent evidence policies, net evidence cost and the perfect-information bound. Keep requested, sent, accepted, executed and outcome-validated actions in distinct states.

Historical quantities and public constructions

[AI] The public numerical constructions are hypothetical and independent of source coordinates. Raw source values, exact conflicts and numerical screens remain in the private annex and code package. Common positive scaling preserves package-weighted ratios and algebraic relations; separately rebasing populations can destroy the reconciliation.

What a future test would establish

[AI] The assigned package comparison would estimate its specified grouping and resource intervention on the eligible original manifest. It would not retrospectively prove the historical routing hypothesis, validate every available bulk-handling option or establish individual responsibility. Transportability requires comparable mechanisms, authority, capacity and outcome measurement.

PRACTICAL HANDOFF

[PD] Recover and reconcile the executed stop lineage, confirm a sanctioned physical response and validate the customer reference. Approve a bounded capacity test before committing scale, causality or financial value.

Private code, evidence maps and hash-bound reviews support reproduction. Historical source verification remains distinct from the reproducibility of a hypothetical calculation.

RESEARCH REFERENCES / GRAIN, ROUTING AND SELECTION

Connect the measured unit to the physical mechanism

Primary research informs the method and prospective design. It does not authenticate the historical source artifacts or establish this project's causal effects.

Kahan, B. C., Li, F., Blette, B., Jairath, V., Copas, A., & Harhay, M. (2023). Informative cluster size in cluster-randomised trials: A case study from the TRIGGER trial. *Clinical Trials*, 20(6), 661-669. doi:10.1177/17407745231186094. Application: distinguish parcel-average from stop-average estimands; fix preassignment population weights. Primary repository abstract, methods/estimand discussion and discussion read; no logistics effect is transferred.

Solomon, M. M. (1987). Algorithms for the Vehicle Routing and Scheduling Problems with Time Window Constraints. *Operations Research*, 35(2), 254-265. doi:10.1287/opre.35.2.254. Application: route feasibility depends on time-window placement and scheduling context, beyond a count threshold. Publisher abstract and metadata accessed; no claim to have reproduced its algorithms or to have read inaccessible full text.

Sigelman, B. H., et al. (2010). Dapper, a Large-Scale Distributed Systems Tracing Infrastructure. Google Technical Report dapper-2010-1. Application: link execution identifiers across service boundaries and document trace sampling. Primary abstract and indexed sections on instrumentation, annotations, sampling and collection read; a trace is not independent proof of physical receipt.

Cole, S. R., et al. (2010; online 2009). Illustrating bias due to conditioning on a collider. *International Journal of Epidemiology*, 39(2), 417-420. doi:10.1093/ije/dyp334. Application: selecting booked losses or approved disputes can distort mechanism comparisons. Full original article read from a university-hosted copy; examples motivate the selection audit, not a claim that its magnitude is known here.

Simpson, E. H. (1951). The Interpretation of Interaction in Contingency Tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 13(2), 238-241. doi:10.1111/j.2517-6161.1951.tb00088.x. Application: marginal and context-specific associations can disagree. Full original article read from a university-hosted copy; the case's flat-curve construction is new and hypothetical.

Kent, K., & Souppaya, M. (2006). Guide to Computer Security Log Management. NIST SP 800-92. doi:10.6028/NIST.SP.800-92. Application: reconcile identifiers, clock consistency, missing logs and retained versions before linking execution events. Official sections 2.3.1-2.3.3 and 3.1 read. Fixed 2006 edition; no current cybersecurity compliance or historical adoption claim.

Ogrinc, G., et al. (2016; online 2015). SQUIRE 2.0 (Standards for Quality Improvement Reporting Excellence): revised publication guidelines from a detailed consensus process. *BMJ Quality & Safety*, 25, 986-992. DOI:10.1136/bmjqs-2015-004411. Indexed primary sections on rationale, context and studying an intervention read; direct publisher and repository access was blocked. Application: distinguish doing improvement from studying whether it caused the result. The case applies a reporting principle across domains; it claims no healthcare effect or formal SQUIRE certification.

Access scopes distinguish full or selected text, primary abstracts, publisher records and indexed sections. Cross-domain methodological applications do not transfer clinical or monetary-policy effects.

RESEARCH REFERENCES / EVALUATION AND UNCERTAINTY

Evaluate original outcomes under shared capacity

Primary research informs the method and prospective design. It does not authenticate the historical source artifacts or establish this project's causal effects.

NIST/SEMATECH. e-Handbook of Statistical Methods, section 6.3.2.2: Individuals Control Charts. Official method. Calculate tentative individual and moving-range limits for ordered source observations. Mixed-grain incident reports, version alternatives and design thresholds are not stable operating phases. A within-limit difference is not evidence of improvement.

Laney, D. B. (2002). Improved Control Charts for Attributes. *Quality Engineering*, 14(4), 531–537. Primary publisher abstract. Dispersion-adjusted rate charts require actual repeated numerator and denominator counts with rational subgroups. A reported approval percentage does not supply its missing exposure. Do not reconstruct fictitious counts from rounded source rates.

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20, 37–46. Primary publisher record. Use a blinded dual-coding sample across approved, rejected, unfiled and unresolved records. Report class prevalence, confusion, kappa and uncertainty. Shared reliance on the same faulty reference can produce agreement without 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. A site selected for extreme defects may subsequently improve without treatment. Quantify regression-to-the-mean sensitivity under explicit reliability assumptions and compare independent periods and nonfocus units; concentration does not identify preventable harm.

Hudgens, M. G., & Halloran, M. E. (2008). Toward Causal Inference with Interference. *Journal of the American Statistical Association*, 103, 832–842. Original manuscript. Shared routes, consignee queues and crews transmit effects between stops. Cluster assignment around the physical capacity domain, retain ordinary work and audit cross-domain sharing. More administrative stops do not create independent experimental units.

Callaway, B., & Sant'Anna, P. H. C. (2021). Difference-in-Differences with Multiple Time Periods. *Journal of Econometrics*, 225, 200–230. Author manuscript. For a staggered observational extension, estimate group-time effects with valid untreated comparisons and stated parallel-trends assumptions. Investigate pre-trends, anticipation and changing mix. A rollout date does not randomize a root-cause mechanism.

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 request through the feasible decisions it changes under stated priors and payoffs. Compare contingent policies, subtract evidence cost and respect the perfect-information bound. Keep this utility separate from service savings.

Access scopes distinguish full or selected text, primary abstracts, publisher records and indexed sections. Cross-domain methodological applications do not transfer clinical or monetary-policy effects.

RESEARCH REFERENCES / CONTRIBUTION AND GOVERNANCE

Keep contribution and value attached to evidence

Primary research informs the method and prospective design. It does not authenticate the historical source artifacts or establish this project's causal effects.

Collier, D. (2011). Understanding Process Tracing. *PS: Political Science & Politics*, 44, 823–830. Primary publication. Grade the specific contribution: nomination, authorship, commissioned scrutiny, sent request, implemented change and measured result. A derivative register permits provisional role evidence; generator metadata and an unobserved original do not independently authenticate authorship.

Tucker, A. L., & Edmondson, A. C. (2003). Why Hospitals Don't Learn from Failures. *California Management Review*, 45(2), 55–72. Author manuscript. Local workaround completion can conceal recurring system conditions. Restore the immediate customer service through authorized work while routing the recurring constraint to its accountable owner. A ticket closure is not evidence that the physical mechanism disappeared.

The W. Edwards Deming Institute. The Funnel Experiment. Official exposition. Adjusting a process to routine variation can increase instability. Separate sparse incident reports from mature exposure-adjusted monitoring before changing grouping rules or announcing a sustained turnaround.

Campbell, D. T. (1976). Assessing the Impact of Planned Social Change. Original institutional record. Accountability can change behavior and its measurement. A stop or dispute metric can improve through labeling and selection while the original customer result is unchanged. Independently follow the original package population.

Goodhart, C. A. E. (1992). The Objectives for, and Conduct of, Monetary Policy in the 1990s. Reserve Bank of Australia Annual Conference. Original author paper, section 1(b)(i). Intermediate measures can lose their relationship to the objective under adaptation. A stop-size target or approval rate needs a verified link to physical customer service; the case transfers the reasoning, not a monetary-policy estimate.

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 strategy, operating actions and value through explicit relationships. A better displayed stop metric or corrected scorecard does not itself demonstrate physical benefit, avoidable cost or cash release.

OECD / European Commission JRC (2008). Handbook on Constructing Composite Indicators. Official handbook. Aggregation requires defined units, denominators, weights and compensation rules. Keep physical service, net economics, ordinary-service protection and authority as distinct requirements. A favorable composite score cannot offset a failed customer safeguard.

Access scopes distinguish full or selected text, primary abstracts, publisher records and indexed sections. Cross-domain methodological applications do not transfer clinical or monetary-policy effects.

FINAL REFEREE REPORT / INTERNAL AGENT-ASSISTED REVIEW

Accept the corrected diagnostic and a falsifiable test

[AI] This author-led, agent-assisted report is not external peer review or client approval. Its disposition concerns the stated evidence boundaries and analytical design.

Major issue	Disposition in the revised case
Unlike counts called inconsistent	Separate incident work, original packages and approved disputes. The missing crosswalk leaves overlap unknown; the arithmetic inequality alone does not prove a discrepancy.
A runtime artifact called decisive	Distinguish effective constraints, realized grouping, shared capacity and customer consequences. One log can resolve one question without establishing the whole causal chain.
A partition-sensitive performance measure	Show unchanged original-package outcomes alongside a changing stop average. Prespecify the package estimand and preserve the original promise.
Flat curve or high approval overinterpreted	Construct the context-mix counterexample; separate selected adjudication from causal responsibility. Approval remains descriptive information within its population.
A grouping fix credited without resource costs	Test the actual service package. Include ordinary-work harm, setup, fully assigned capacity and valid outcome truth; economics and information value remain separate.

Partner critique: can a COO act on page one?

[AI] Yes: verify the executed service unit and customer outcome, preserve fair scorecard review and require a feasible physical response before committing to a root cause or remedy. The contribution is specific: diagnostic authorship, scope judgment and commissioned scrutiny, provisionally supported by the register.

Remaining conditions and minor issues

[AI] Original recovery, effective policy, population links, service times, field capacity, reference accuracy and trial recruitment remain calibration needs. The illustrative model does not establish those inputs. Reproducibility is not external validation or a consulting-firm ranking; historical dispatch, adoption and savings remain unverified.

REVIEW DISPOSITION / PASS WITHIN STATED SCOPE

[AI] Accept the corrected diagnostic and prospective evaluation. Preserve the unresolved attribution question. A failed authority, customer, measurement, economic or ordinary-service condition prevents expansion under the proposed decision rule.

Internal source, methodological, numerical, executive and actual-page review; no external referee endorsement or verified client implementation is claimed.