

OPERATING CASE STUDY / CUSTOMER EXPERIENCE & LOGISTICS

Delivery Promise Integrity

Convert a contested service-window finding into a testable operating decision

An anonymized operating case on linking delivery-window grades to concession records, withdrawing an unsupported reporting claim, and designing a capacity decision around the original customer promise.

PROJECT CONTEXT / DELIVERY WINDOWS, MEASUREMENT & CAPACITY DECISIONS

BLUF / BOTTOM LINE UP FRONT

Validate the promise and the comparison.
Keep every original order in the test.

ANALYTICAL FINDING

Identical queries do not guarantee an unbiased ratio. A missing comparison category cannot test a reporting gap. A gate can improve the accepted-order score while leaving the original customers no better served.

CONTRIBUTION IN THE SOURCE RECORD

Authored the analysis, documented limits, withdrew an unsupported claim and routed decisions. Provisional registry hoop support.

EVIDENCE STATUS

Originals unrecovered. Executive delivery was rejected; filter validation, operational actions and realized benefit remain unestablished.

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

EXECUTIVE ABSTRACT / DECISION AND OWNERSHIP

Make the promise measurable before committing capacity

[AI/PD] The Promise Integrity Contract (PIC) connects the customer’s original commitment to a valid physical outcome, a reconciled financial record and an operating test on all original demand. A large recorded load contrast prioritizes investigation; it does not identify a preventable effect.

Abstract element	Decision-relevant synthesis
Context [HR]	A same-day delivery analysis joined internal slot grades to concession records. It reported a large load contrast, withdrew a proposed reporting-gap claim and proposed a capacity review before stations accepted cycles.
Objective [PD]	Determine which commitments were actually missed, whether the comparison survives data repair, and whether a capacity-allocation policy improves original-customer outcomes at acceptable resource cost.
Approach [RC/AI]	Reconcile promise versions, units and distinct packages, group-specific contamination, financial maturity, missing scan outcomes and the absent day-level comparison category. Separate a diagnostic finding from a treatment effect.
Findings [AI]	The same query can reverse a comparison when contamination differs. An immature ratio is not necessarily a lower bound. A better accepted-cycle score can coexist with unchanged failure across original orders.
Decision value [PD]	Fund a bounded reconciliation first. Advance to a calibrated capacity-allocation test only after original promises and outcomes can be verified, with rejected and rerouted demand retained in the denominator.
Implications [PD]	Data owners validate the comparison; customer-policy owners resolve promise classes; Operations specifies the actual capacity action; Finance validates resource costs. The evaluator tests original-order outcomes and displacement.

EXECUTIVE RELEASE CONDITION

[PD] Preserve the original customer window, the end-of-day commitment where applicable, all booked demand and existing entitlements. Require independent physical outcomes and a valid economic threshold before scaling the gate.

Basis: source register S1. [HR] historical record; [RC] recalculated; [AI] analytical inference; [PD] proposed design. No historical causal improvement, implemented capacity gate or realized saving is established.

EXHIBIT 1 / PROMISE AND MEASUREMENT

An internal expedited target is not automatically the customer’s promise

[HR] The source distinguishes internal slot-level grading, a separate external day-level view and an unresolved free-upgrade versus genuine-window-miss split. It also records a correction: operational teams already used slot-level reporting. [AI] The gap concerns alignment and linkage, not proven organizational blindness.

KEEP FOUR RECORDS LINKED WITHOUT MAKING THEM INTERCHANGEABLE

Proposed promise contract / classify the original communicated commitment, not whether an upgrade was paid



A free upgrade may be a customer commitment if communicated. Preserve the actual promise.

Basis: PIC design. Internal grades, customer contracts, receipt and accounting answer different questions.

Contract layer [PD]	Required record and interpretation
Customer commitment	Preserve the window or day the customer was actually promised at booking, its timezone and later versions. A free internal speed upgrade does not automatically replace that original contract.
Internal operating target	Retain the assigned slot, dispatch plan and service class. Missing an internal target can be operationally relevant even when the original customer promise is met.
Physical outcome	Use independently verified receipt time and nonreceipt status. An attempted delivery or missing scan is not automatically equivalent to successful receipt or permanent loss.
Financial response	Link distinct concession events, eligible unit quantities and posting dates. A financial response can reflect several defects, policy rules or delayed customer contact.

ALIGN THE CONSTRUCT BEFORE RANKING THE DRIVER

[PD] Pair original promises with actual receipt, then compare those outcomes with internal and external grades. Resolve the policy distinction without rewriting past promises or treating every internal miss as customer harm.

Basis: S1 recorded measurement-policy question and correction [HR]; PIC paired promise/outcome contract [PD]. The actual external grade and customer-promise mapping were not reproduced.

EXHIBIT 2 / RELATIVE LOAD VALIDITY

The same summation rule can reverse the intended comparison

[HR] The source acknowledged that selecting packages with a target concession could bring other concession units into the sum. It treated the ratio as robust because both groups used the same definition. [AI] That conclusion requires equal proportional contamination, not merely identical code.

IDENTICAL CODE DOES NOT IMPLY EQUAL PROPORTIONAL CONTAMINATION

Wholly hypothetical / 100 packages in each group; target and other concession units are distinct

MISSED SLOT

Target units: 30

Other units: 90

Recorded total: 120

ON-TIME SLOT

Target units: 40

Other units: 0

Recorded total: 40

Recorded ratio = 120 / 40 = 3.00 Target-only ratio = 30 / 40 = 0.75

Basis: original counterexample. Common definitions cancel error only under additional symmetry conditions.

[RC] In the hypothetical exhibit, each group has 100 packages. The missed-slot group has 30 target units plus 90 other units; the on-time group has 40 target units and no other units. The recorded load ratio is 120/40 = 3, while the target-only ratio is 30/40 = 0.75. The same selection-and-sum rule operates in both groups.

[AI] Concession units per package are not a binary customer-concession probability. Multiple units, multiple events, repeated joins and non-target categories can each alter the numerator. A distinct package count does not deduplicate financial events or establish a causal comparison.

Validation gate [PD]	Evidence that changes the decision
Category and grain	Reconcile target-only units using event identifiers, package identifiers and category rules. Compare each group’s all-record sum with its deduplicated target-only sum.
Filter sensitivity	Replay the ship-option and service-class filters on identical original populations. Preserve unmatched packages; establish group-specific coverage rather than assuming symmetric omissions.
Maturity sensitivity	Read matched promise cohorts at comparable ages. A partially observed ratio can be above or below its mature value when the two groups mature differently.

A LARGE CONTRAST IS A RECONCILIATION PRIORITY

[AI] Keep the association as a source-reported finding pending validation. Do not translate it into a customer probability, a causal effect, a guaranteed lower bound or a return on a capacity investment.

Basis: S1 declared unit-sum caveat and unclosed filter challenge [HR]; original counterexample [RC]. All exhibit quantities are invented and disclose no historical measurement.

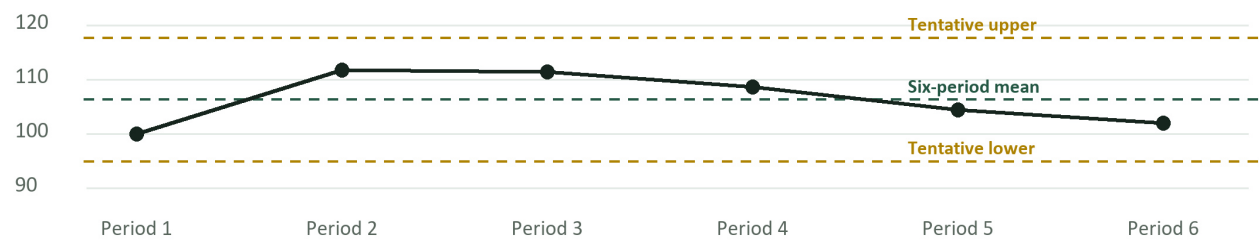
EXHIBIT 3 / TENTATIVE PROCESS BEHAVIOR

A short sequence does not identify capacity as the cause

[HR] The source interpreted repeated slot misses as a structural problem and proposed a capacity-and-difficulty gate. It did not measure capacity over-commitment by site or cycle, and incident mechanisms remained unresolved. [RC/AI] A process screen cannot supply that missing causal evidence.

NO POINT EXCEEDS THE TENTATIVE REFITTED INDIVIDUALS BAND

Historical recorded graded-miss rates / first period = 100; raw counts and transformation remain private



Six points cannot establish stability or identify capacity, routing or incidents as the cause.
Basis: S1 weekly graded counts and tentative XmR arithmetic. Event completeness remains an evidence requirement.

[RC] The chart uses count-derived graded-miss rates, with the first period indexed to 100. Individuals limits use the mean plus or minus 2.66 times the average adjacent moving range. No point exceeds the tentative refitted band. With six observations, that is a descriptive screen; it does not establish a stable process or rule out incidents.

[AI] The final event grade need not wait for concession posting, but it still depends on complete and correctly interpreted events. Missing scans, service-class mix, late-arriving records and changed filters can affect the event denominator. An immature financial cohort remains unsuitable for an unqualified load comparison.

Interpretation [AI/PD]	Decision consequence
Within tentative limits	Do not call each move improvement or deterioration. Investigate mechanisms using site, cycle, route and promise-class evidence.
Capacity is a hypothesis	Compare committed work, effective capacity, dispatch constraints and routing difficulty before choosing reserve capacity as the intervention.
Exceptions can coexist	A persistent aggregate level can contain episodic failures and offsetting subgroup movements. Separate those paths rather than forcing a single explanation.

Basis: S1 weekly graded counts [HR]; count-derived first-period-100 XmR screen [RC]. Raw values, rounding alternatives and all other process screens remain in the private annex.

EXHIBIT 4 / MISSING OUTCOMES

Exclude an unobserved order and the metric changes before service does

[HR] The reported slot-miss denominator excluded orders with no same-day scan. Their outcome truth and concession loading were not fully reconciled. [AI] Missing observation is a separate state; it cannot be silently counted as success or automatically equated with nonreceipt.

MEASURE ALL ORIGINAL ORDERS AND SHOW WHAT REMAINS UNKNOWN
Wholly hypothetical / 100 original orders: 40 known misses, 50 known successes, 10 unresolved



Basis: missing-outcome bounds, not imputed historical results. All unknowns remain in the denominator.

[RC] In this invented cohort, 40 known misses and 50 known successes produce a graded-only miss rate of 40/90 = 44.4%. With 10 unresolved original orders, the full-cohort miss risk is bounded by 40% and 50%. Neither endpoint is an estimated historical rate.

Preserve in the ledger [PD]	Why it matters
Every original booked order	An unsuccessful induction, refusal, diversion or missing join must not erase a customer from the treatment population.
Reason for unresolved truth	Distinguish absent scan, delayed ingestion, missing identity, unobserved delivery and confirmed nonreceipt. These states imply different repairs.
Outcome bounds and recovery	Report known outcomes over the original denominator. Audit unresolved records and recalculate bounds as evidence arrives, without using treatment-induced observability as eligibility.

A COVERAGE IMPROVEMENT IS NOT YET A SERVICE IMPROVEMENT
[PD] Track measurement completeness and physical outcomes separately. A successful scan recovery may correct the score without changing the delivery; a real recovery may help the customer without immediately changing the financial record.

Basis: source-recorded excluded no-scan population [HR]; hypothetical partial-identification example [RC]; original-demand follow-up design [PD]. No missing historical record is imputed as a failure.

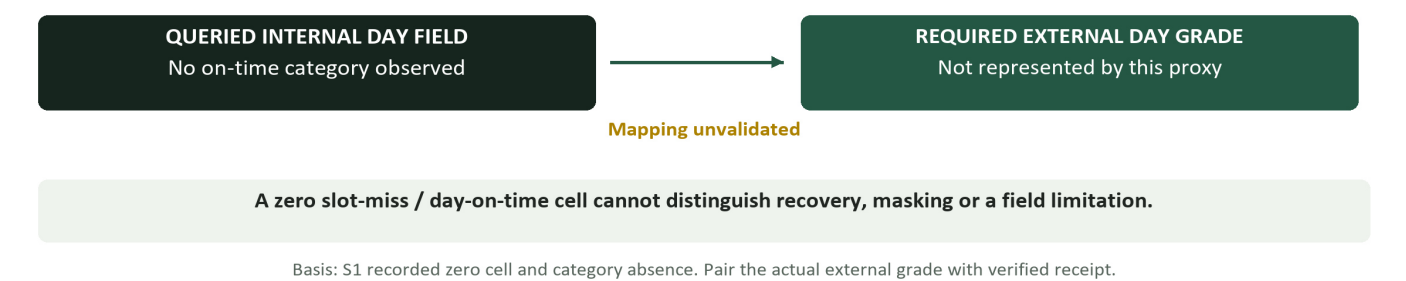
EXHIBIT 5 / DIAGNOSTIC TEST VALIDITY

A zero cell cannot test a category the table never records

[HR] The proposed reporting-gap cross-tab returned no cases with an internal slot miss and a day-level on-time grade. The source also records that the day-level on-time category never appeared anywhere in the queried cut. [AI] This result does not discriminate between competing external-reporting explanations.

ABSENCE OF A CATEGORY IS NOT A TEST OF EXTERNAL REPORTING

Schematic of the source-described support failure / no raw count is disclosed



[AI] A null result is informative only when the measurement can express the event being tested. Here the proxy field’s support was the problem. The observed absence does not establish that slot misses never recover, that the external view never masks them, or that the external view necessarily does.

Required paired test [PD]	Prespecified observation
Validate the positive category	Recover known examples of externally on-time orders and determine whether the queried field represents that classification at all.
Build the correct cross-tab	Link original customer promise, independent receipt, internal slot grade and actual external grade on the same entities and maturity horizon.
Preserve unknowns and classes	Separate genuine customer-window commitments from internal free upgrades; retain unmatched records and test classification reliability independently.

WITHDRAW THE CLAIM; KEEP THE QUESTION OPEN

[HR/AI] The author recorded the negative result and withdrew the unsupported masking claim. That is defensible analytical restraint. The unresolved mapping still needs evidence and a policy owner; the zero itself does not settle it.

Basis: S1 cross-tab and absent day-level category [HR]; diagnostic-support inference [AI]; new paired-validation protocol [PD]. The source description of a falsification is narrowed to the inference the table permits.

EXHIBIT 6 / CAPACITY GATE DESIGN

An accepted-order score can improve while original customers do not

[HR] A capacity review before a station accepted a cycle was proposed but not shown to be implemented. [AI] Its evaluation must retain all previously booked orders. Otherwise the gate can remove difficult demand from the measured population without improving physical service.

FILTERING DIFFICULT ORDERS CAN IMPROVE THE SCORE WITHOUT IMPROVING SERVICE

Wholly hypothetical measurement-only scenario / physical delivery outcomes remain fixed

ORIGINAL ORDERS 40 misses / 100 orders 40%	EXCLUDE 20 MISSED ORDERS 20 misses / 80 accepted 25% reported	FOLLOW ORIGINAL CUSTOMERS 40 misses / 100 original 40% unchanged
---	--	---

A real gate must redeploy capacity and improve outcomes across the original demand.

Basis: denominator-selection counterexample. No historical manipulation or gate outcome is alleged.

[RC] The invented measurement-only example starts with 100 original orders and 40 misses. Excluding 20 missed orders leaves 20 misses among 80 accepted orders: the reported rate falls from 40% to 25%. All 40 original customers still miss. This demonstrates a denominator mechanism, not misconduct or a historical effect.

Gate requirement [PD]	Evaluation consequence
Original commitment fixed	Record booked orders and their original promises before the operating offer. The gate may redeploy or reroute capacity; it may not earn success by widening promises.
All demand followed	Retain refused, rerouted and unserved orders. Measure original-slot receipt and end-of-day failure over the original eligible population.
Displacement fully charged	Include reserve effort, coordination, failed allocation and any delay moved to another slot, route or customer. Shared capacity determines the randomization domain.

MAKE THE TREATMENT PHYSICAL AND THE DENOMINATOR FIXED

[PD] The prospective model tests forecast-guided reserve allocation within approved constraints. A classification change, acceptance filter or narrower denominator is not the operating benefit.

Basis: S1 unimplemented capacity-gate proposal [HR]; original selection counterexample [RC]; PIC test design [PD]. Customer promises and safety constraints remain fixed.

EXHIBIT 7 / DECISION FIT AND CONTRIBUTION

A reproducible association still has to answer the commissioned decision

[HR] The record attributes the query, finding note and reproducibility appendices to the author. It also records rejection by the commissioning reviewer: a relative load contrast did not answer the requested network-impact decomposition. The replacement view and operational actions were not established as complete.

ASSOCIATION, NETWORK CONTRIBUTION AND PREVENTION ARE DIFFERENT CLAIMS

Decision-fit ladder / each step requires evidence that the preceding step cannot supply



Reproducible arithmetic does not guarantee relevance to the executive question.
A completed analysis does not establish acceptance, implementation or realized benefit.

Basis: source-recorded rejected executive deliverable, unclosed sensitivity and unimplemented capacity proposal.

Claim [HR/AI]	Contribution grade and release boundary
Analysis and publication	Provisional registry hoop support for authoring the query, documenting limits and producing the analysis. Unrecovered originals prevent independent authentication of the full result table or code.
Correction and escalation	Provisional registry hoop support for withdrawing the masking claim, accepting a filter threat and routing operational and policy decisions. Disclosure is not closure of the sensitivity test.
Operating and financial effect	Not established. No implemented gate, post-action outcome contrast or realized saving is evidenced. The author did not own the whole delivery program or external measurement policy.

[AI] Reconstruct descriptive network contribution only after target units, overlap and the common denominator reconcile. A subgroup rate is not additive to a network rate; a sibling outcome category is not an input to the target category. Converting an association into a preventable network effect requires a causal estimand and an effective intervention.

NEXT DECISION: VALIDATE, THEN TEST

[PD] Release the paired promise and data audit first. Advance a capacity test only when the comparison, original cohort, independent outcomes, costs and recruitment are credible. Provide the sponsor’s required decision unit alongside the analytical finding.

Basis: S1 authoring, role boundaries, open sensitivities and rejected delivery [HR]; bounded process tracing [AI]. Internal peer checks do not establish client acceptance or an operational result.

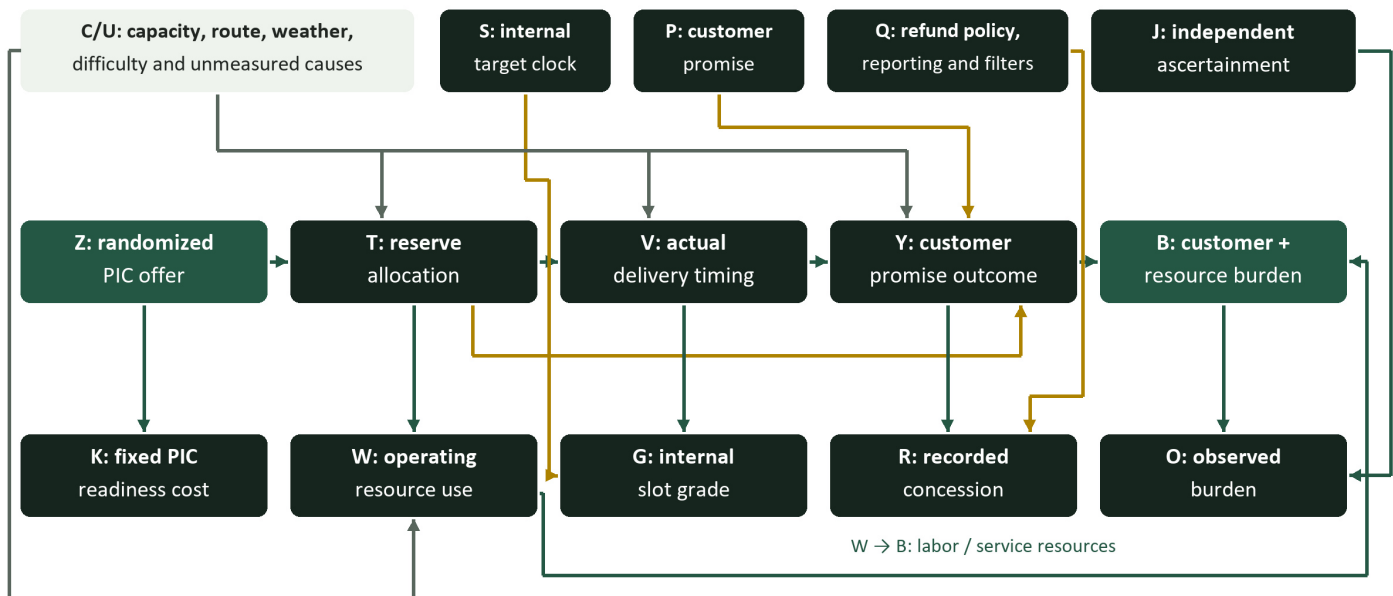
TECHNICAL ANALYSIS

Separate the promise, the delivery and the financial record

[AI/PD] The proposed Promise Integrity Contract (PIC) links an original customer commitment to physical receipt, then evaluates a capacity response. Internal grades and concessions remain distinct measurement channels.

VALIDATE THE CUSTOMER PROMISE; TEST THE OPERATING RESPONSE

Proposed PIC graph / original orders and promise versions are frozen before assignment



C/U → T, V, Y, W. The direct T → Y path allows other handling or displacement effects.

Y → R ← Q separates physical service from concessions. G does not replace Y.

K is charged once outside B. All original orders remain; missing truth stays unresolved.

Crossing lines without junction dots add no edges. Arrows are hypotheses, not historical effects.

[PD] Z assigns PIC; T allocates the existing reserve; V is actual delivery timing. The original communicated promise P and V determine customer outcome Y. Internal target S instead defines grade G. A free upgrade can become a customer commitment if communicated; payment is not the classifier.

[AI/PD] Y and operating resources W produce burden B, observed through independent ascertainment J as O. Fixed PIC cost K is charged outside B once. Recorded concessions R also depend on policy, reporting and filters Q. C/U includes capacity, route, weather, difficulty and unmeasured causes.

[AI] The direct T→Y path retains possible handling or displacement effects beyond the planning model. Conditioning on available grades, accepted cycles or linked concessions can select the cohort. All original orders remain in the proposed effect. These arrows specify hypotheses; the register identifies no intervention coefficient.

Basis: registered promise, grade and concession distinctions [HR]; proposed PIC causal structure [AI/PD]. Physical outcomes and administrative records are not interchangeable.

TECHNICAL ANALYSIS

Define the contract before deciding whether a delivery failed

[HR/AI] The record compares an internal slot outcome with concession-unit load. It does not authenticate the original customer promise for every package or identify the effect of changing execution.

Estimand [AI/PD]	Unit, population and interpretation
Internal graded miss share	Missed internal grades / missed plus on-time grades, in one frozen source version. Describes the graded population only.
Customer promise-breach risk	Original eligible orders that fail their communicated window or day promise / all original eligible orders, with unknown truth bounded.
Concession-unit load ratio	Mean eligible target concession units per package in one slot group divided by the corresponding mean in the other group. Units may be fractional or multiple.
PIC offer effect	Usual-minus-offer mature customer/resource burden over all originally assigned orders; fixed implementation cost charged separately once.

[AI] A customer window contained within a promised calendar day imposes a stricter timing requirement when both use the same receipt event, time zone and eligibility. Delivery outside that window can still satisfy the day promise. That logical possibility does not establish its frequency in the historical data. Missing scans do not establish either success or failure.

[PD] Store the communicated commitment, acceptance time, original window boundaries, later authorized changes and the internal operating target separately. A tighter uncommunicated target can be an internal failure without breaking a day-only promise. Conversely, a communicated upgrade can create a real customer obligation without an extra fee.

[HR/AI] The registered join sums units on packages carrying a target concession flag. This is not a binary package-concession probability. Daily unit counts divided by same-day miss counts also mix units, cohorts and reporting clocks; they cannot establish absolute package conversion.

[PD] Prespecify original-promise follow-up, source version, terminal receipt categories and financial-event eligibility. Do not repair a failed original promise by moving its deadline after assignment. Report the later agreed service separately if a customer authorizes a change.

Basis: registered field definitions and unvalidated promise-policy split [HR]; explicit descriptive and causal estimands [AI/PD]. The proposed trial has no historical result.

TECHNICAL ANALYSIS

The same query does not make unequal errors cancel

[AI/RC] A shared calculation can preserve a bias instead of removing it. The claimed robustness of the relative concession load needs group-specific validation of target units, joins and reporting maturity.

Independent hypothetical example	Slot-missed group	Slot-on-time group
Original packages	100	100
Intended target concession units	30	40
Additional non-target units included	90	0
Published load under the same rule	120 / 100	40 / 100
Result	Recorded ratio 3	Target ratio 0.75

[AI/RC] Let $L_m, L_o > 0$ be true target-unit loads and C_m, C_o the extra loads included. Then $R_{\text{observed}}/R_{\text{target}} = (1 + C_m/L_m)/(1 + C_o/L_o)$. The ratio is unchanged exactly when proportional contamination is equal. Applying identical SQL does not impose that condition. The hypothetical table reverses the direction without changing the rule between groups.

[PD] Reconcile target-bucket units before the package join. Audit multiplicity, fractional allocation, reissued identifiers and group-specific filter losses. A distinct-package denominator does not undo duplicated financial units. The requested with/without service-filter rerun addresses one threat; it does not alone establish correct units or customer-promise identity.

[AI/RC] Maturity creates another ratio condition. For fixed cohorts and nonnegative accumulating loads, let $F_m(H), F_o(H) > 0$ be the fractions captured by horizon H . Then $R_H = R_{\text{final}} \times F_m(H)/F_o(H)$. A ratio is a floor only when $F_m(H) \leq F_o(H)$. A final ratio of 2 appears as 4 when capture is 0.8 versus 0.4, and as 1 when those fractions reverse.

[HR/AI] Thus the immature historical ratio is not automatically conservative. Obtain group-specific revision curves; allow reversals and changing policy rather than assuming pure accumulation. The source association remains a hypothesis-generating observation, not a verified causal or monetary multiplier.

Basis: exact independent hypothetical constructions [AI/RC], reproduced privately. Differential-error theory and reporting-delay methods motivate the audit; no example estimates historical contamination.

TECHNICAL ANALYSIS

Unknown outcomes and unsupported categories require different tests

[HR/AI] The registered slot-miss denominator excludes ungraded records. Separately, the internal day-status field never supplies its on-time category in the reported cut. Neither fact can be resolved by naming every unknown a failure.

Independent hypothetical cohort	Calculation	Meaning
100 original; 90 graded; 40 misses	40 / 90 = 44.4%	Conditional graded miss share
All unknown outcomes succeed	40 / 100 = 40%	Original-cohort lower bound
All unknown outcomes fail	50 / 100 = 50%	Original-cohort upper bound

[AI/RC] These sharp bounds require mutually exclusive, exhaustive and correctly linked original entities. They do not estimate which unknowns failed. Historical identifiers counted distinctly within several cells are not necessarily distinct across the complete extract. Validate cross-cell exclusivity before summing population totals.

[HR/AI] A zero slot-missed/day-on-time cell is uninformative about real recovery when the day-on-time category is absent throughout that source. The recorded decision to withdraw the masking assertion was appropriate; the query did not establish that external day reporting has no masking or that slot misses never recover. A valid falsification test first needs a measure capable of expressing the predicted outcome.

[PD] Audit a random core of 200 future order-context records, supplemented by separately sampled missing-grade, upgrade, boundary-time and join-exception records. Blinded coders reconstruct the communicated promise and receipt from independent evidence. Under independent binary observations, 200 gives a worst-case 95% agreement precision screen of about ±6.9 percentage points; shared source/cycle clustering requires recalculation.

[PD] Retain first-pass confusion matrices, category prevalence, raw agreement and nominal kappa intervals; kappa precision requires pilot simulation. Apply sampling weights before population extrapolation. Adjudicate disagreements without erasing unresolved truth. Agreement on the same faulty field is not external validity.

Basis: registered missing-grade and unsupported day-status categories [HR]; exact synthetic bounds [RC]; proposed reliability audit [PD], informed by Cohen. Historical unknowns are not imputed.

TECHNICAL ANALYSIS

Association, reporting lag and policy do not identify the remedy

[HR/AI] The original analysis redirected attention toward slot execution and away from upstream freight. A strong conditional association can justify investigation, but neither comparison identifies which operating intervention prevents customer harm.

[AI] Capacity, route complexity, weather, incident severity and original promise type can influence both slot success and concessions. Refund rules can change concessions without changing receipt. Selecting linked, graded or accepted-cycle records can add bias. Compare fixed original populations and retain preassignment covariates; do not control away post-treatment grades when estimating total PIC effects.

[AI] A different freight exposure with a small crude association is not automatically a negative control. It can plausibly affect service and need not share the same confounding or measurement paths. Negative-control theory requires those assumptions explicitly. The sibling analysis therefore neither eliminates upstream causality nor leaves slot execution as the only possible driver.

[HR/AI] Weekly aggregate regressions and same-day incident comparisons address different quantities from package-level load. They may be weakened by aggregation, changing exposure or reporting delay, but those explanations are untested here. Reporting delay does not make contemporaneous correlation mathematically impossible: persistent operating shocks and mixed event/report cohorts can create it. A lag mismatch neither validates the larger association nor disproves the alternative.

[AI/RC] Selection can mimic operating success. In an independent hypothetical cohort of 100 orders with 40 misses, removing 20 difficult misses from reporting gives $20/80=25\%$, while the unchanged original-cohort miss rate remains 40%. A capacity gate that declines or reroutes orders needs access and customer outcomes over original demand, not only the accepted-volume rate.

[PD] Keep the mechanism test concrete: does better placement of an existing reserve improve verified customer service under unchanged promises and capacity? Decompose refusals, other-slot displacement, later receipt and nonreceipt. A statistical association between an internal signal and concessions does not validate that signal as a substitute for the intervention outcome.

Basis: registered observational comparisons and untested mechanism [HR]; negative-control and surrogate-endpoint theory [AI]. The selection example is invented and exact, not a historical estimate.

TECHNICAL ANALYSIS

Repeated levels do not diagnose a stable capacity constraint

[HR/AI] A narrow band of reported miss rates prompted a structural-capacity explanation. Repetition establishes that similar values were reported; it does not isolate capacity, rule out incidents or establish a stable measurement process.

[AI/RC] For eligible ordered individuals, compute $MR_t = |x_t - x_{(t-1)}|$, center $\text{mean}(x)$, individual limits $\text{mean}(x) \pm 2.66 \text{ mean}(MR)$, and MR upper limit $3.267 \text{ mean}(MR)$. Preserve source order and maturity; do not refit an incomplete endpoint into a baseline and call the result stable.

[RC/AI] The private inventory contains 33 screens: 11 event/count descriptions, eight load/maturity screens, three output/zero screens, three sparse contextual endpoints and eight mechanical version/display/target comparisons. The count-derived graded miss series has no tentative X or MR crossing. Some missing-observation and immature-load endpoints cross their arithmetic reference limits. These are conditional measurement alerts, not identified operating changes.

[RC/AI] Exact transcribed numerator/exposure pairs permit a conditional six-period rate calculation if package states are unique binary outcomes and cohorts are valid. The ordinary binomial chart flags every subgroup; a Laney adjustment does not. This disagreement reflects substantial dispersion relative to independent Bernoulli sampling. Neither method authenticates the process, and the record establishes no stable eligible operational panel.

[PD] Investigate shared cycle failures, changing mix, repeated identifiers, grade capture and maturity before interpreting overdispersion. Do not reconstruct counts from rounded rates or pool fractional concession-unit loads into a binomial model. A constant zero masking series has no moving-range information about a field that cannot express the positive category.

[AI/PD] No observed worst-N intervention cohort permits an empirical regression-to-the-mean correction. For future site selection, freeze eligibility, estimate untreated ranking reliability and use a concurrent comparison. Under the hypothetical linear relation $E[\text{next}-\mu | \text{baseline}] = p(\text{baseline}-\mu)$, excess d naturally shrinks by $(1-p)d$; vary p rather than calling it measured.

[PD] Prespecify promise type, cycle, business line, node type, route density and baseline capacity as moderators. Test interaction contrasts with intervals. Do not infer different effects merely because one subgroup is significant and another is not.

Basis: all provided numeric series and private conditional SPC calculations [RC]; NIST/Laney and selection methods. Raw levels and complementary historical indices remain private.

TECHNICAL ANALYSIS

The defensible contribution is an auditable question, not proven impact

[HR/AI] The register attributes the query specification, association analysis, reproduction appendices and withdrawal of an unsupported masking assertion to the analyst. The record also preserves peer challenge and a commissioning rejection. These support a bounded analytical contribution.

[HR/AI] Original query files, the complete raw result and underlying discussion have not been recovered for this review. Byline and session descriptions are registry transcriptions, so contribution remains provisional process-tracing hoop evidence. Independently authenticated scripts, run outputs and reviewer receipt would strengthen the chain; they would not by themselves identify an intervention effect.

[HR] The proposed operational gate, measurement-policy decision and source-filter repair belonged to other accountable roles. Closure and post-intervention customer outcomes are not established in the supplied corpus. A later dashboard listing does not prove ownership, adoption or causal benefit. The absence of recovered closure is not proof that no work happened elsewhere.

[AI] Agatz and colleagues make time-window design a service/cost tradeoff. Campbell–Savelsbergh studies customer-accepted choices, which differ from retrospectively changing an accepted commitment. Measurement-error theory and Lawless’s reporting-delay work explain why a shared query or cutoff cannot make data valid automatically. The proposed PIC unites these constraints around original customer demand.

[AI] Deming makes the missing stable measurement basis a decision concern. Campbell/Goodhart makes excluding difficult demand a concrete target distortion. An OECD-style composite cannot compensate for absent promise identity with a strong concession association. Ittner–Larcker–Randall cautions against assuming additional measures create value; Tucker–Edmondson motivates learning from the reviewer’s unmet decision need. These are scoped interpretive connections, not borrowed effect estimates.

[AI] Consolidated limits are original-source access, customer/internal promise alignment, unknown grades, cross-cell identity, differential contamination, reporting maturity, unrun filter sensitivity, observational confounding, missing capacity evidence and unmeasured implementation. The requested conversion to impact requires an identified intervention contrast and costs; multiplying the observed load difference by volume would not supply either.

Basis: derivative role/action record [HR]; bounded process tracing and literature synthesis [AI]. No downstream ownership, financial realization or sibling-project outcome is assigned to this analysis.

TECHNICAL ANALYSIS

Test reserve placement against the same original customer commitments

[PD] The new prospective PIC evaluation randomizes independent capacity domains to a pre-cycle warning and reserve-allocation protocol or usual allocation. Both receive the same reserve capacity and approved authority; every original order remains in scope.

[PD] Freeze six two-slot cycles, their externally scheduled original orders and communicated promises before assignment. The synthetic law uses customer-window versus genuine day-only contract probabilities, not payment status. An internal earlier target alone does not rewrite the customer contract. A warning uses information available before the cycle; the latent future bottleneck is not granted to operators.

[PD] One existing reserve crew targets one of the two slots. Better targeting can recover potential misses, while orders in the other slot still require monitoring. The model explicitly distinguishes later same-day receipt, later-day receipt and nonreceipt at original promise plus 28 days. Unknown reference truth is retained through bounds, never silently scored as success.

[PD] The primary ITT compares usual-minus-offer customer/resource burden per original assigned order. Customer-promise failure is separately reported. Common crew resource and incremental PIC readiness belong to the same horizon, with readiness charged once. Costs are hypothetical normative resource units; the historical concession association supplies no efficacy, forecast accuracy or valuation parameter.

[PD] Randomize domains wide enough to contain shared staff and inter-slot displacement, preserve parallel assignment and record cross-domain interference. Post-assignment grading changes, reserve uptake and accepted-cycle volume are mediators or diagnostics, not eligibility filters for the total effect. Later-day harm, handling safety, refused access and oracle error need separate thresholds and power; the primary calculation does not establish those safeguards.

[PD] Use an independent pilot to validate warnings, physical receipt, dependence and feasible recruitment. Compare detectability at economic break-even with joint customer and economic acceptance. The model pages state a conditional design and stress tests, not an executed trial or available field sample. Re-plan if the required independent domains or reference quality cannot be obtained.

Basis: new PIC trial protocol [PD], developed independently of historical associations. All probabilities, costs and power inputs in the planning model are explicitly hypothetical.

TECHNICAL ANALYSIS

Four release propositions prevent a strong ratio from becoming a weak decision

[AI/PD] PIC is a proposed contract with disconfirming tests. A reproducible statistic, a useful operating intervention and a verified customer benefit require different evidence.

Proposition [PD]	Acceptance and disconfirmation
P1: Promise identity is valid	Independent records reproduce communicated windows, internal targets and receipt. Unsupported categories or unresolved versions block a promise-breach claim.
P2: Relative load survives measurement repair	Group-specific target-unit, join, filter and maturity audits preserve the claimed comparison within a preregistered tolerance. Otherwise restate or withdraw it.
P3: PIC improves original-demand service	The randomized all-order customer contrast clears its prespecified threshold without changing original commitments or excluding difficult demand.
P4: Full-cost value clears the decision	Economic confidence bound clears readiness cost, while separately specified safety, access and measurement gates pass. A favorable point estimate alone fails.

[PD] Preregister meaningful effects, horizon, analysis weights, baseline covariates and missing-truth bounds. Require adequate precision for equivalence or noninferiority margins; nonsignificance is not proof of no benefit or harm. Approval belongs to the operating and customer-policy owners, not to the analyst’s choice of a convenient statistic.

[PD] A later nonrandomized staggered rollout should estimate group-time ATT against comparable not-yet-treated domains. Assess no anticipation, common support, plausible parallel trends, original-promise invariance and spillover. Event-study and equivalence diagnostics can challenge assumptions, not prove them. Mature outcome and covariance pilots must precede field power or savings claims.

[AI] Information value may justify the targeted source and promise audit before an operating trial. Its payoff model is separate from customer benefit and cannot be added to the same avoided loss. A decision to gather evidence cannot override an unresolved mandatory safety, authority or measurement gate.

[AI] The immediate recommendation is to authenticate the customer contract and repair the comparison before treating it as an investment case. Keep the source’s useful analytical challenge while retiring its unsupported ratio-cancellation, causal-elimination and maturity-floor claims.

Basis: proposed four-part PIC framework [PD]; causal, financial and measurement limits [AI]. A future observational ATT and the randomized planning ITT are distinct quantities.

PROSPECTIVE MODEL / FIXED PROMISES AND RESERVE CAPACITY

Test where to deploy reserve capacity, not which customers to exclude

[PD] This hypothetical trial randomizes independent shared-capacity domains to forecast-guided reserve allocation or usual early-slot allocation. Both receive one reserve crew per cycle. Preserve all externally scheduled orders and original communicated customer contracts before assignment.

Frozen element	Invented planning law
Population and follow-up	Six cycles per domain; 120 orders/cycle, 60 per window. Verify physical receipt at original promise +28 days. Rejected, rerouted and unobserved orders remain.
Customer commitment	Window promise probability 0.80; day-only promise 0.20 with an uncommunicated earlier internal target. Types are drawn before assignment; counts are not fixed. Free communicated upgrades may create commitments.
Capacity mechanism	Of 24 potential internal-window misses, 20 are in the bottleneck slot and 4 in the other. Reserve recovers Binomial(20,0.60) when correctly placed or Binomial(4,0.50) otherwise.

[PD] Domain grades have early-slot bottleneck probabilities 0.30/0.70 and forecast accuracy 0.70/0.90, equally likely. Forecasts use only pre-cycle information. Conditional cycles are independent within a persistent grade; domains are independent. The crew cannot serve both slots.

THE CUSTOMER CONTRACT DEFINES FAILURE

[PD] Remaining internal misses resolve 75% later the same day, 20% later-day receipt and 5% nonreceipt. Same-day receipt breaches a genuine window contract but can meet a day-only contract. Concession triggers are secondary records; no refund or customer entitlement changes.

[PD] All parameters are hypothetical, not fitted to the reported association. The 28-day readout requires calibration. Aggregate distributional ordering does not establish that every individual order benefits.

PROSPECTIVE MODEL / FINITE DECISION POWER

Detect the economic threshold before funding a large-effect story

[PD] Finite cycle distributions are mixed by shared domain grade and convolved across independent domains. Planning assumes this probability law is known; actual pilot uncertainty requires recalibration. A favorable full-effect result does not guarantee adequate power near break-even.

Design	Domains / arm	MDE (index)	Break-even detection	Joint power floor
Initial	32	11.97	17.8%	77.5%
Base candidate	224	4.59	80.9%	>99.9%
Same / stress	224	5.79	59.8%	87.6%
Revised / stress	384	4.45	83.4%	98.5%

[PD] Usual variable burden is 100; fixed incremental cost is 4.64 index points. Require 80% correct-direction power at the positive 2.5% tail of a two-sided 5% zero-effect test, and an 80% joint floor for economic and original-customer-promise superiority. Economic clearance uses a one-sided 97.5% lower bound.

[PD] The response path mixes usual allocation with the forecast policy. Stress widens bottleneck probabilities to 0.15/0.85 and forecast accuracies to 0.52/0.88. The 224-domain candidate retains 87.6% joint power at its assumed effect but only 59.8% detection at break-even; it fails.

A NUMERICAL REDESIGN IS NOT A RECRUITMENT COMMITMENT

[PD] The 384-per-arm stress candidate requires 768 independent capacity domains and 552,960 original orders. No minimal allocation or available recruitment is claimed. Separately size day-breach, rushing, safety, off-domain displacement and missing-reference guardrails before any trial proceeds.

[PD] Joint floor = max(0, economic power + promise power – 1); no endpoint independence is assumed. Inclusive finite-tail inversion and monotone aggregate loss distributions calibrate the economic boundary under the stated law.

PROSPECTIVE MODEL / FULL RESOURCE ECONOMICS

Value original-contract outcomes and charge the gate in full

[PD] Hypothetical resource units value a genuine same-day window miss at 3, later-day receipt at 9 and nonreceipt at 20. An uncommunicated internal miss that meets the customer’s day promise receives no customer-harm charge. Operational target misses remain separately visible.

[PD] Common reserve expense 8/cycle is included in both arm levels. Usual variable burden 86.2 is indexed to 100. PIC adds 4/cycle for amortized forecast setup/monitoring 1.5, coordination 1.5 and production verification 1.0. Research-only truth collection is separate; these are not historical dollars.

Avoidable share a	h=0.80	h=0.90	h=1.00
0.60	0.48	1.12	1.76
0.80	2.19	3.04	3.90
1.00	3.90	4.97	6.03

[PD] Cells are stress net index points; h is the non-overlap share. $Net_raw = a \times h \times 9.2 - K$; $Net_index = 100 \times Net_raw / 86.2$. Discount only avoided harm; new costs remain full. At full realization, $K=2/4/6$ gives net 8.35/6.03/3.71 index points. Reprice categories separately if realization differs.

[PD] If half of nonreceipts later become later-day receipts, the stress benefit falls from 9.20 to 8.65 raw units and net falls to 5.39 index points. That is another maturity estimand requiring new power. Missing truth remains bounded on original denominators.

NO CUSTOMER EFFECT IS ESTIMATED BY THE SOURCE RATIO

[PD] Field value requires original eligible volume × mature randomized outcome-vector ITT benefit at avoidable, nonoverlapping prices, less all new work and adverse effects. Customer response, operating interference, prices and recruitment remain uncalibrated. A separate observational extension requires its own identified estimand.

[PD] Amounts are per original 120-order cycle; divide by 120 for a per-order valuation. Common crew cost cancels only in the contrast. Gate cost is allocated to the same pool regardless of research sample size.

PROSPECTIVE MODEL / SEPARATE VALUE OF EVIDENCE

Distinguish a capacity opportunity from a promise or trigger problem

[PD] This separate decision model first isolates insufficient evidence or unsafe capacity; otherwise it distinguishes a supported physical capacity opportunity from a measurement or concession-trigger issue. Priors, signals and utilities are hypothetical.

Next investment action	Physical capacity issue	Measurement / trigger issue	Insufficient evidence
Bounded capacity-allocation trial	14	-8	-15
Reconcile promises, joins and triggers	3	10	-2
Hold investment; acquire evidence	1	1	6

[PD] Priors are 0.45/0.30/0.25. The best prior action has utility 3.85; perfect information can add at most 6.95. Signal accuracy q assigns each wrong signal probability $(1 - q) / 2$. All 27 signal-to-action policies are evaluated.

Signal quality q	EVSI	Evidence cost	Net EVSI
0.50	0.4125	0.40	0.0125
0.70	3.0275	1.40	1.6275
0.90	5.6425	3.30	2.3425

EVIDENCE VALUE IS NOT ANOTHER OPERATING SAVING

[PD] Evidence can favor a bounded trial, a data repair or an investment hold. Acquisition costs are separate from downstream action costs already included in utility. Do not add EVSI to the operating model. Authority, customer promises and safety gates remain mandatory regardless of expected utility.

[PD] This Bayesian decision exercise does not authenticate the recorded ratio, imply a measured causal effect, or authorize automated acceptance, promise changes or customer compensation restrictions.

Reproduce the comparison before accepting the intervention story

[AI] This case documents an analytical contribution and its limits. Its new contribution is a testable contract that preserves the actual customer promise, distinguishes physical outcomes from financial records and follows original demand through a capacity decision.

Claim class	Meaning and replication boundary
[HR] Historical record	A statement or action attributed by the supplied register. Original SQL, the full result table, correspondence and the authored document remain unrecovered. Recorded bylines and confidence labels do not independently authenticate them.
[RC] Recalculated	Count-derived rates, rounded-load reconciliation, tentative process limits and constructive counterexamples. Three transcribed example rows do not reconstruct the full source dataset. Arithmetic does not establish causality.
[AI] Analytical inference	Conditional interpretations: unequal contamination can reverse a ratio; unequal capture can reverse a maturity bound; an absent comparison category cannot resolve external reporting; selective acceptance can improve a score without service.
[PD] Proposed design	The Promise Integrity Contract, paired-grade validation, original-demand capacity-allocation study, resource sensitivity and decision-information model. These are new designs, not historical implementations.

Replication sequence

[PD] Freeze the source and selected cells. Reconcile the weekly counts and each rounded load version, preserving the distinction between a weighted display calculation and the unrecovered raw sum. Keep every unresolved financial and physical outcome visible.

[PD] Audit communicated promises, event identifiers, category-specific units, filters and observation ages by group. Validate the actual external-grade mapping. Then independently enumerate the proposed finite operating law, calibrate its assumptions and verify power and recruitment before a trial.

THE DECISION THE CASE CHANGES

[AI/PD] Move from a dramatic association to an evidence-gated choice: repair the comparison, validate the promise, and test reserve allocation on original demand. Neither a cleaner accepted-order denominator nor a delayed financial response proves prevention.

Only one historical miss-rate series is indexed to its first period. Raw measurements, identities, source paths, adjacent-project data and the transformation key remain private.

RESEARCH REFERENCES / DELIVERY WINDOWS AND MEASUREMENT

Test the promise that was made and the action that can change delivery

Primary research and official methods inform the measurement contract and proposed evaluation. They do not authenticate the unrecovered project originals or establish a historical capacity-gate effect.

Agatz, N. A. H., Campbell, A. M., Fleischmann, M., & Savelsbergh, M. W. P. (2011). Time Slot Management in Attended Home Delivery. *Transportation Science*, 45, 435–449. DOI 10.1287/trsc.1100.0346. Publisher abstract and metadata. Delivery-window design trades service availability against operating cost. This case distinguishes designing a future offer from executing a previously communicated promise.

Campbell, A. M., & Savelsbergh, M. (2006). Incentive Schemes for Attended Home Delivery Services. *Transportation Science*, 40, 327–341. DOI 10.1287/trsc.1050.0136. Full published author copy. The study models incentives affecting customer window choices. It does not justify retrospectively widening an existing commitment; original-customer service remains visible in the proposed experiment.

VanderWeele, T. J., & Hernán, M. A. (2012). Results on Differential and Dependent Measurement Error of the Exposure and the Outcome Using Signed Directed Acyclic Graphs. *American Journal of Epidemiology*, 175, 1303–1310. DOI 10.1093/aje/kwr458. Primary indexed abstract/introduction; direct access challenged. Shared measurement machinery does not establish nondifferential error. The case's contamination proof is independently derived.

Lawless, J. F. (1994). Adjustments for reporting delays and the prediction of occurred but not reported events. *Canadian Journal of Statistics*, 22, 15–31. Publisher DOI presentation 10.2307/3315826.n1. Publisher abstract and metadata. Reporting-delay variability motivates explicit ascertainment models. No universal maturity deadline or ratio lower bound follows from delayed posting alone.

Lipsitch, M., Tchetgen Tchetgen, E., & Cohen, T. (2010). Negative Controls: A Tool for Detecting Confounding and Bias in Observational Studies. *Epidemiology*, 21, 383–388. DOI 10.1097/EDE.0b013e3181d61eeb. Primary indexed full text. A negative control needs an appropriate causal and bias structure. A weak crude association for another operational exposure does not rule out that exposure's causal role.

Prentice, R. L. (1989). Surrogate endpoints in clinical trials: Definition and operational criteria. *Statistics in Medicine*, 8, 431–440. DOI 10.1002/sim.4780080407. Publisher abstract and metadata. Association alone does not establish that changing an indicator transmits an intervention effect. A verified customer-window breach is an outcome in its own right; an internal speed target may be a different construct.

Access scope is recorded in the private research memo. Methods and operating analogies do not import another study's effect size, commercial result or certification.

RESEARCH REFERENCES / UNCERTAINTY AND EVALUATION

Let the evaluation survive missing observations and shared capacity

Primary research and official methods inform the measurement contract and proposed evaluation. They do not authenticate the unrecovered project originals or establish a historical capacity-gate effect.

NIST/SEMATECH. e-Handbook of Statistical Methods, section 6.3.2.2: Individuals Control Charts. Official method. Process limits need comparable ordered observations. The case reports short-series calculations and frozen-reference sensitivities without treating every movement as a signal or inferring the cause of an excursion.

Laney, D. B. (2002). Improved Control Charts for Attributes. *Quality Engineering*, 14(4), 531–537. Primary publisher abstract. The case computes a conditional dispersion-adjusted chart from supplied count pairs. Large denominators do not validate disjoint classification, event completeness or an operating phase; estimated dispersion does not reveal its cause.

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20, 37–46. Primary publisher record. Independent coding, confusion cells and uncertainty support reliability assessment. Agreement on an internal grade cannot itself validate the original customer promise or external metric.

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 adverse sites can distort later comparisons. Freeze eligibility and retain concurrent controls; aggregate source counts do not identify an empirical regression-to-the-mean correction.

Hudgens, M. G., & Halloran, M. E. (2008). Toward Causal Inference With Interference. *Journal of the American Statistical Association*, 103, 832–842. Primary manuscript. Shared crews and capacity create spillover. The proposed domain boundary must contain material dependencies, and residual displacement requires observation beyond the nominally treated slot.

Callaway, B., & Sant'Anna, P. H. C. (2021). Difference-in-Differences with Multiple Time Periods. *Journal of Econometrics*, 225, 200–230. Author manuscript. A later nonrandom staggered rollout needs group-time effects, eligible comparisons and explicit assumptions. It cannot inherit identification or power from the hypothetical randomized planning example.

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. Information has value through the choices it changes. The case separately values investigation and its cost, without adding that value to the same operating benefit twice.

Access scope is recorded in the private research memo. Methods and operating analogies do not import another study's effect size, commercial result or certification.

RESEARCH REFERENCES / OPERATING LEARNING AND CONTRIBUTION

Reward a valid customer outcome rather than a convenient score

Primary research and official methods inform the measurement contract and proposed evaluation. They do not authenticate the unrecovered project originals or establish a historical capacity-gate effect.

Collier, D. (2011). Understanding Process Tracing. *PS: Political Science & Politics*, 44, 823–830. Primary publication. Discriminating evidence supports contribution claims. Query authorship, documented correction, executive acceptance, implemented action and resulting benefit are distinct claims; the register provides provisional support for only the stated roles.

Tucker, A. L., & Edmondson, A. C. (2003). Why Hospitals Don't Learn from Failures. *California Management Review*, 45(2), 55–72. Author manuscript. Local responses can leave system causes unresolved. A rejected analytical deliverable motivates a clear decision contract and follow-through; no healthcare effect is transferred to this case.

The W. Edwards Deming Institute. The Funnel Experiment. Official exposition. Reacting to routine variation can increase instability. This case separates tentative process screens from a demonstrated capacity mechanism and requires evidence before choosing an operating adjustment.

Campbell, D. T. (1976). Assessing the Impact of Planned Social Change. Original institutional record. Consequential indicators can alter behavior and recording. The original-demand denominator prevents an acceptance rule from earning performance credit merely by removing difficult orders. This is a design safeguard, not a claim of historical intent.

Goodhart, C. A. E. Problems of Monetary Management: The U.K. Experience. 1975 conference; revised volume listed as 1976 in the RBA archive. Targeting can change a measure's relationship to the intended outcome. Freezing communicated promises distinguishes genuine service gains from revised commitments or selective reporting.

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. Measurement must connect to strategic use. A striking load comparison and reproducible calculation do not automatically answer a sponsor's network-impact question.

OECD / European Commission JRC (2008). Handbook on Constructing Composite Indicators. Official handbook. Weighting, normalization and aggregation affect meaning. The proposed burden measure exposes physical outcomes and resource costs separately; a net score cannot silently hide later delivery, nonreceipt or displaced work.

Access scope is recorded in the private research memo. Methods and operating analogies do not import another study's effect size, commercial result or certification.

FINAL REFEREE REPORT / INTERNAL AGENT-ASSISTED REVIEW

Accept the analytical contribution within its measurement boundaries

[AI] This is author-led, agent-assisted review, not external peer review, client approval or a consulting-firm ranking. Its verdict concerns source-bounded interpretation and a falsifiable proposed study; it certifies no historical prevention or financial benefit.

Major issue	Disposition in the revised case
Load versus probability	Preserved units per package as a load statistic. Differential contamination, filters and joins can alter the ratio despite identical code; source confidence does not close these tests.
Promise and missing outcomes	Separated original communicated promises from internal upgrades, retained unknown outcomes and preserved original demand through capacity decisions.
Zero and maturity claims	An absent day-level positive class cannot test external masking. Unequal financial capture makes the immature ratio a possible overstatement or understatement.
Mechanism and evaluation	Did not treat a short series or sibling crude association as proof of capacity failure. Specified a new physical reserve-allocation test with explicit dependence, resource and recruitment assumptions.
Contribution and relevance	Credited authorship, correction and escalation provisionally. Retained the rejected commissioned delivery, open filter test, unimplemented gate and absence of a realized result.

Partner critique: can a COO act on page one?

[AI] Yes: require a paired promise/outcome audit and a deduplicated target-unit comparison before funding a capacity test. The sponsor can ask for the correct decision unit while retaining the author’s documented diagnostic and correction work.

Minor revisions and remaining limits

[AI] Recover originals and complete the unrun sensitivities. Calibrate customer-promise classes, forecast performance, outcome ascertainment, resource weights, dependencies and feasible recruitment before using the proposed planning result.

REVIEW DISPOSITION / PASS WITHIN STATED SCOPE

[AI] The case supports analytical leadership and the proposed Promise Integrity Contract. It establishes no validated external-reporting gap, historical capacity-gate effect or realized saving.

Scope: internal source, scientific, numerical and executive review. Acceptance is conditional on the stated evidence boundaries; no consulting-firm rank or external endorsement is implied.