

OPERATING CASE STUDY / CUSTOMER EXPERIENCE & LOGISTICS

Network Flow & Customer Promise Protection

A gate can improve its own metric while delaying the original customer

An integrated operating-programme case that separates stage coverage, physical control, shared service capacity and mature outcomes. It tests whether an added check earns its time and cost.

FLOW CONTROL & ORIGINAL-CUSTOMER SERVICE

BLUF / BOTTOM LINE UP FRONT

Test the control on the original promise. Keep every delayed parcel in the result.

ANALYTICAL FINDING

Preventing a recorded process defect does not guarantee timely service. A control can repair the parcel it checks while delaying other parcels. Stage coverage and a portfolio haircut cannot identify that trade-off.

CONTRIBUTION IN THE SOURCE RECORD

Provisional registry-level hoop for framing an integrated programme and its decision narrative. Delivery and realized impact remain separate.

EVIDENCE STATUS

Exact plan and build originals remain unrecovered. A related planning workbook does not authenticate acceptance or execution.

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

STRUCTURED ABSTRACT

Require a customer-service test before committing to a control architecture

[AI] A process can become harder to bypass while the original customer becomes more likely to wait. Control effectiveness and service preservation must be evaluated together.

Element	Decision-relevant conclusion
Context [HR]	A planning lead reportedly assembled an integrated reduction proposal spanning physical controls, software and feedback. The record describes a proposed architecture and its funding narrative, not authenticated deployment or measured customer improvement.
Objective [AI]	Determine what each control changes, whether it can operate within shared capacity, and how its incremental value survives original-customer measurement and cheaper feasible alternatives.
Approach [RC/AI]	Reconcile programme arithmetic privately; separate stage coverage from joint response. Test the effect of added service time on original deadlines using an independent finite construction. Preserve uncertainty about adoption, causal effects and resource availability.
Findings [AI]	A gate may eliminate recorded defects among dispatched parcels while increasing original-deadline failures. Coverage, an overlap haircut and selected sensitivity scenarios do not establish a guaranteed reduction or minimum system size.
Decision value [PD]	Operations defines lawful controls and capacity; the customer owner fixes promises and remedies; Finance validates avoidable costs; an independent measurement owner retains delayed and unresolved parcels. Compare a bounded check with a cheaper bounded alternative before commitment.
Implications [PD]	Use the Flow-and-Service Evidence Contract (FSEC). The invented replay preserves a cheaper-rule advantage and inadequate economic precision. A staffed redesign has positive expected value under its assumptions but still does not justify a launch.

[PD] Begin with dependency and duration validation, then an authorized test on original cohorts. Do not infer a production decision from a generated plan, a clean process metric or this synthetic benchmark.

[HR] historical record; [RC] recalculated; [AI] analytical inference; [PD] proposed design. Two derivative records and a related planning artifact do not authenticate the original plan, approval or execution. Historical coordinates remain private.

EXHIBIT 1 / COVERAGE AND RESPONSE

A complete stage map does not establish a complete causal explanation

[AI] Locating controls along a process is useful design work. It does not establish their reach, reliability, joint effect or necessity.

Architecture claim	Evidence needed before using it
Every stage has a programme	Original exposure, exceptions and actual control execution at each stage.
A hard gate prevents the failure	False passes, false holds, bypasses, missing signals and downstream consequences.
Each programme adds protection	Joint response under shared units and a common continuation, including interactions.
Every layer is indispensable	Dependency-specific removal tests; unique coverage alone cannot establish complementarity.
Feedback makes the system improve	Independent labels, complete sampling, drift checks and governed model releases.

[HR/AI] The source links several programmes to a smaller set of stages. It is not a one-programme-per-stage experiment, and no such matching would prove the architecture optimal. A defect originating upstream can be detected or repaired downstream. An origin share therefore does not establish an interception ceiling for a later control.

[AI] Pure overlap and complementarity are different mechanisms. An overlap adjustment can prevent counting the same affected units twice. It cannot demonstrate that removing any layer destroys the system. Conversely, a prerequisite with little standalone value can be valuable in a bundle. The claim needs an explicit joint response, not a label or a uniform haircut.

TEST A DEPENDENCY-COMPLETE INCREMENT

[PD] Specify a lawful subset, its prerequisites and an independent outcome. Preserve a no-added-control and a simpler alternative. A small test need not omit an essential dependency; an all-at-once programme need not be the smallest valid experiment.

[HR/AI] The source architecture is proposed. Stage attribution, programme estimates and external precedents are not measured effects for this operating population.

EXHIBIT 2 / THE ORIGINAL DEADLINE

A perfect dispatched-unit metric can conceal worse customer service

[PD] An independent FIFO example exposes the denominator problem without assuming a queueing distribution or borrowing a historical service rate.

Invented ten-day system	Usual flow	All-unit gate
Original arrivals / same-day promises	9 per day / 90 total	Same 90 original parcels
Daily service capacity	10	8
Original-deadline failures	6 intrinsic failures	54 queue-related failures
At the end of day ten	No queue backlog	44 delivered late; 10 pending
Selected completed-unit process defect	Not the customer endpoint	0 among 80 handled

[PD] Arrivals occur at each day's start and FIFO service finishes at day end. The gate repairs the intrinsic defects among handled parcels but consumes capacity. Daily cohort deadline failures become 1,2,3,4,5,6,7,8,9,9. The ten pending parcels are only a lower bound: previously completed late parcels still missed their original promises.

[AI] Excluding held parcels from dispatch eligibility can make the selected process measure look perfect. Returning, reassigning or later rescuing a parcel does not erase the original promise. Those actions may still be valuable, but their benefit belongs in a separately timed outcome. The original cohort must retain them through the comparison.

USE PHYSICAL RECEIPT, NOT A CLEAN RELEASE LIST

[PD] Count original-deadline receipt, later recovery and unresolved outcomes separately. Compare unchanged eligibility and equivalent ascertainment. Do not let a status update, itinerary removal or different defect label become evidence that the customer received the parcel.

[PD] Wholly invented deterministic construction, independently recalculated. It is not a forecast of the source programme or a claim that every gate necessarily harms service.

EXHIBIT 3 / TIME, CAPACITY AND SAFE EXECUTION

Fit the check inside the service window before starting it

[AI] An observation window, a congestion threshold and an incremental service time are different quantities. Their units cannot be exchanged to justify a launch.

Quantity	Operational question
Sensor observation window	How long is evidence associated with an event, and does collection block service?
Physical occupancy threshold	What space or exposure is measured, with which denominator and response?
Added check time	How much work does each attempt consume, including retries and unsuccessful checks?
Remaining original slack	Can the authorized check finish before the unchanged dispatch/customer deadline?
Shared resource capacity	Which other parcels, routes or tasks lose service when the control runs?

[HR/AI] The source includes sensing, hold and congestion mechanisms. A sensor's matching window is not automatically time added to every parcel. A physical occupancy threshold is not queue utilization. Inferring stability or delay from those substitutions would add an unsupported causal mechanism to the plan.

[PD] A bounded controller starts a check only when its validated maximum duration fits the remaining slack. It sees current slack and completed results, never future check duration or latent physical truth. If duration can exceed the bound, the protocol needs an approved interruption or fallback, with all consumed time, unresolved work and spillover retained.

CALIBRATE THE BOUND, NOT ONLY THE AVERAGE

[PD] Measure tails, failed attempts, concurrency, setup and resource displacement. Reserve capacity explicitly and pay for it even when unused. Zero ordinary harm in a model with a perfect duration bound is a structural assumption, not a field safety finding.

[PD] No historical sensing interval or congestion threshold calibrates the proposed replay. Existing customer remedies and authorized status controls remain unchanged.

EXHIBIT 4 / DEPENDENCIES AND FEEDBACK

A dependency graph needs time and resource evidence to become executable

[AI] Logical precedence identifies what must come first. It does not alone establish a critical path, completion date or available team.

Planning element	Release evidence
Prerequisite edge	Accepted output, interface version and failure route.
Task or programme owner	Accepted authority and available capacity, distinct from a name in a table.
Delivery sequence	Durations, resource conflicts, integration tests and readiness dates.
Feedback loop	Independent labels, time-indexed learning and deployment versions.
Stage or pilot completion	Verified execution before mature customer and resource outcomes.

[HR/AI] The plan describes staged deployment and feedback. These are proposed schedules, not completed releases. A feedback cycle cannot be interpreted as a one-time acyclic delivery plan without separating learning and release periods. A stated iteration frequency does not establish that the labels are correct or the next version improves outcomes.

[AI] Training only on the defects remaining in a selected downstream feed can omit held, rerouted, relabeled and delayed parcels. The policy can change who is observed. Keep an independent sample from the original population, including non-alerted and unresolved cases, so the learning system does not reward its own exclusions.

MAKE EACH RELEASE REVERSIBLE AND TESTABLE

[PD] Freeze the action and measurement versions, dependency acceptance, resource envelope, stop criteria and rollback owner. Preserve the lineage from the original parcel through every reassignment. Validate execution and outcome separately; a completed task does not establish a successful intervention.

[HR/AI] Named programmes, owners and waves establish planning intent in the selected records. Acceptance, resource commitments and delivery remain separate evidence states.

EXHIBIT 5 / ECONOMIC AND SENSITIVITY LIMITS

A reduction estimate is not yet an avoidable-cost or funding case

[AI] Counts, rates, resource costs, benefit entitlement and cash release require different bridges. Arithmetic closure does not supply missing definitions.

Claim to reconcile	Required common basis
Programme reductions to portfolio	Unique original units, joint response and a stated continuation.
Rate movement to avoided failures	Same exposure denominator, metric definition and observation horizon.
Avoided failures to value	Applicable unit costs, realization, overlap and retained harm.
Benefit estimate to investment case	Incremental build, operation, maintenance and displaced-work costs.
Sensitivity scenario to a lower bound	A defined assumption set that includes every claimed scenario.

[RC/AI] The private reconciliation distinguishes rounded totals, selected programme reductions, deferred components and ambiguous cost-versus-entitlement labels. Selected cases do not establish a universal floor. A lower-bounded baseline is not a fixed baseline; annualizing a selected weekly count is not a mature annual forecast. Adjacent versions require a period and population bridge before being called contradictory.

[PD] Retain three distinct economic states: value not yet verified, verified value not yet avoidable, and avoidable cost not yet released as cash. Unknown value is not automatically zero. Equally, a desired reduction or a repeated headline cannot be treated as realized benefit. Preserve unfavorable scenarios and a cheaper feasible operating alternative.

DO NOT DISCOUNT THE HARM WITH THE BENEFIT

[PD] Apply realization and nonduplication only to favorable changes under an explicit comparison. Charge extra delay, failed checks, ordinary-flow displacement, paid standby and reference work in full. Cost a redesigned operating package before asking whether more observations would help.

[RC] Historical arithmetic and all raw quantities remain private. [PD] The model prices original-deadline failures and actual check work in invented resource units, not historical dollars.

EXHIBIT 6 / CONTRIBUTION AND DECISION AUTHORITY

Credit the architecture without borrowing the programmes' outcomes

[AI] The supported contribution concerns framing and decision preparation. It does not inherit execution, effects or credentials from the programmes discussed.

Activity or evidence	Bounded conclusion
Integrated programme narrative	Provisional registry-level hoop for the planning lead's analytical contribution.
Cited operational analyses	Inputs with their own authors, clocks, definitions and validation limits.
Generated plan or script description	Reported artifact production; exact originals remain unrecovered.
Reported funding or carryover disposition	A typed source claim; a decision record is needed to authenticate acceptance.
Related planning artifact	Evidence of that artifact, not proof the original system was implemented.
Customer improvement and savings	Require independent mature outcomes and an appropriate causal and economic comparison.

[HR/AI] The records describe a coherent planning effort that connects failure stages, control concepts, resources and a funding narrative. They also describe reasons for not advancing the entire system and later reuse of selected ideas. Those explanations are not an identified causal account of a funding decision or evidence that the controls delivered their estimated effect.

[AI] A named accountable person is not automatically the author of every input or the owner of operational deployment. A related workbook cannot authenticate an unrecovered original or its handoff. Conversely, an unrecovered file does not prove no discussion, implementation or benefit occurred elsewhere.

KEEP FOUR CLAIMS SEPARATE

[PD] Authored analysis, authenticated submission, authorized deployment and realized outcome each need their own evidence. Treat the present case as internal analytical assurance of a bounded record and proposed test, not an external endorsement or verified programme success.

[HR/AI] Exact plan, generator and supporting original analyses were not recovered in the bounded search. A related planning workbook does not resolve acceptance or execution.

EXHIBIT 7 / FLOW-AND-SERVICE EVIDENCE CONTRACT

Choose the smallest control that earns its time and cost

[PD] FSEC links the original promise to a validated service bound, an authorized action and a mature all-cohort comparison. It is a release discipline, not proof of benefit.

FSEC obligation	Decision record
Freeze the original customer frame	Original parcels, deadlines, eligibility, remedies and later recovery separately.
Validate control and time	Physical confirmation, false results, maximum duration and concurrency.
Reserve and price shared resources	All attempts, standby, failed work and displacement, including unused capacity.
Retain cheaper feasible alternatives	No added check, a single bounded check and a slack-aware rule.
Test mature outcomes and economics	Original-deadline receipt, ordinary-flow harm, full resources and uncertainty.

[PD] In the invented base replay, two bounded checks improve expected service versus no added check but have less net resource value than a cheaper one-check rule. Under tighter slack they have negative net resource value. Staffing changes the operating package and improves its mean, yet the selected design remains too imprecise to clear economic break-even.

[PD] Perfect knowledge of hidden workload grade cannot improve the base resource decision: the same cheap action remains best in each grade. More information therefore cannot rescue that purchase under the locked objective. A service-priority objective could select differently, but its willingness to pay and acceptable harm must be agreed explicitly.

CALIBRATE BEFORE COMMITTING

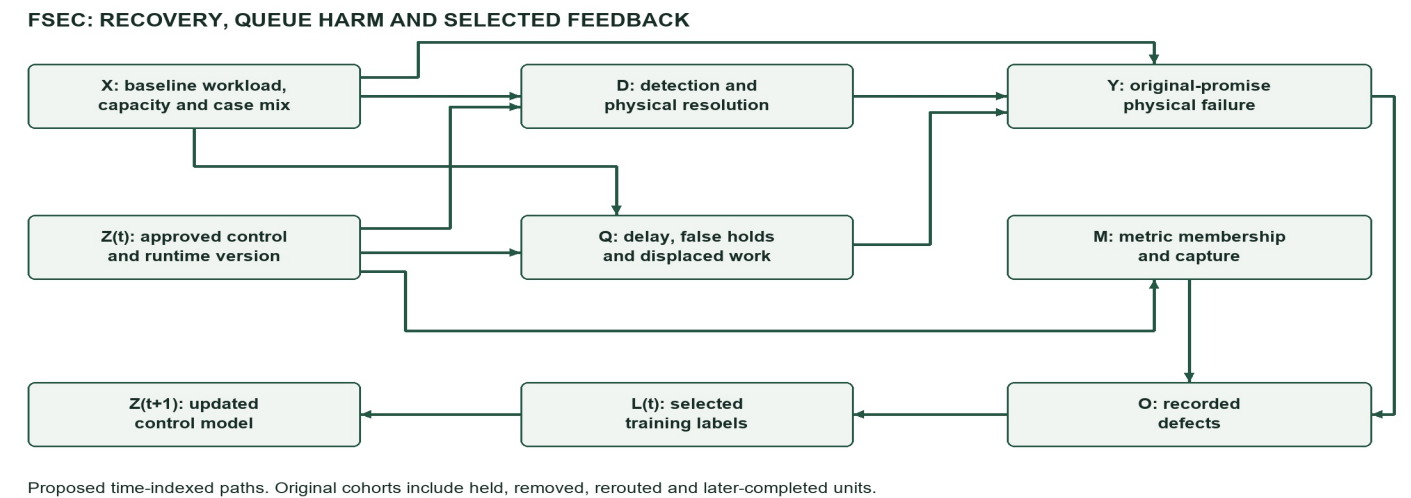
[PD] Validate time tails and independent receipt measurement first. Prespecify the comparator, prices, service threshold and resource assumptions. Authorize an operating experiment only after feasibility and safety review; none of the model's proposed allocations is presented as an achieved or approved field test.

[PD] Entirely hypothetical replay. Point benefit, statistical detectability, economic clearance and field feasibility are separate decisions. No public historical quantity calibrates the model.

TECHNICAL ANALYSIS

Bind control performance to end-to-end service

[PD] The Flow-and-Service Evidence Contract (FSEC) binds the original eligible cohort, physical recovery, permitted control actions, queue capacity and the original customer deadline.



[PD] Z assigns the approved control; D is detection and physical resolution; Q is queue delay and false holding; Y is original-promise failure. Z can improve D while worsening Q. X contains preassignment workload, mix and capacity. M is metric membership and capture; O is recorded defects. Training labels L at time t influence the next model version Z at t+1. The time index prevents a feedback loop from masquerading as a static DAG.

[HR/AI] A programme-to-stage map does not prove one control per stage, exhaustive causes or a joint causal reduction. Multiple controls may address distinct failure paths at the same stage. An upstream origin does not prevent downstream interception; a downstream sensor does not guarantee correct identity, authority or timely recovery. The proposed shares and overlap adjustment are modeled assertions, not observed causal effects.

[HR/AI] The registry describes preparation and an architecture; it does not authenticate system implementation or a realized reduction. Exact plan and build originals were not recovered. A related planning artifact cannot authenticate this plan, its acceptance or execution. Repeated byline, script and narrative references remain one provenance lineage. Contribution is a provisional registry-level hoop for framing and synthesis, not sole authorship or operating benefit.

Basis: source interpretation [HR/AI] and proposed causal graph [PD]. Prentice (1989); Bareinboim/Pearl (2016); Perdomo et al. (2020).

TECHNICAL ANALYSIS

Perfect detection can coexist with worse service

[AI] Independent finite construction: nine bookings arrive at the start of each of ten days. Every original promise expires at that day’s end. FIFO service carries unfinished work forward.

Hypothetical process	Usual	Verification gate
Daily handling capacity	10	8
Original bookings	90	90
Intrinsic failures after processing	6	0
Completed by final cutoff	90	80
Completed after original deadline	0	44
Still pending at cutoff	0	10
Unique original-deadline failures	6	54
Recorded intrinsic defects / completed units	6 / 90	0 / 80

[AI] The six usual intrinsic failures are the first six bookings on day one. The gate detects and repairs all six before their deadlines. Its additional work nevertheless reduces daily capacity. Under FIFO, the number from each arriving cohort missing that day’s deadline is 1, 2, 3, 4, 5, 6, 7, 8, 9 and 9. These sum to 54 distinct original bookings: 44 eventually processed late and ten still pending.

[AI] Conservation alone gives a weaker bound: 90 arrivals minus 80 completions leaves at least ten original deadline failures. Ten is the final backlog, not the total ever late. Exact 54 requires the stated FIFO order and daily deadlines. A different scheduling rule changes that count but cannot make all 90 timely with only 80 completions.

[AI] The reported defect measure excludes pending units and counts only intrinsic misplacement among processed units. It improves to zero while original-promise failures increase. This is not a claim that virtual removal or verification is inherently harmful: either can enable timely rescue. The necessary evidence is the original-cohort physical outcome plus the induced queue and cost.

[PD] Restoring capacity to nine per day removes queue lateness in this fixture. Staffing, changed sequencing or a bounded check policy is a different treatment whose reserved capacity must be priced. A larger trial cannot repair an adverse operating mean. No stationary queue or historical calibration is assumed. Spearman and colleagues motivate evaluating flow, backlog and service together; the exact counts here are newly derived.

Basis: independently enumerated FIFO proof [AI], wholly hypothetical. All historical quantities remain private. Spearman et al. (1990).

TECHNICAL ANALYSIS

Measure missed service outside the surviving label pool

[PD] Freeze original booking identity, eligibility and deadline before any hold, removal, reroute or dispatch decision. Keep every route and unit in its assigned denominator.

[HR/AI] A scan, motion signal, sealed container or vehicle-assignment event is evidence about a process, not an infallible physical receipt reference. Missing evidence is not proof no event occurred. Event time, recording time, original promise and eventual recovery are different clocks. A later recovery does not erase a missed original deadline. Source-stated input correlations and foreign precursors do not establish local causal effects.

[PD] Independently reconstruct receipt by the original promise, with ascertainment at original promise plus 28 days in the proposed benchmark. Validate the late-capture and reopening tail before field use. Sample dispatched, held, virtually removed, rerouted and unresolved units with recorded inclusion probabilities. Report original-cohort missing-reference bounds; do not silently exclude hard cases or change the outcome to eventual receipt.

[PD] Two coders blinded to assignment and operational verdict assess identity, timing, physical resolution and failure category. Report weighted confusion, prevalence, raw agreement and cluster-bootstrap kappa; adjudicate against independent artifacts. Four hundred independent binary assessments give worst-case simple-agreement precision near plus/minus 4.9 percentage points, not kappa power or clustered precision. Check invariance across device version, parcel mix, station type, workload and baseline slack.

[AI/PD] Lakkaraju and colleagues show why decision-selected labels can mislead evaluation. A model trained only on remaining recorded defects can miss harm to held or removed units. Perdomo and colleagues distinguish stable retraining from optimal performance when deployed rules change the data. Freeze training and evaluation vintages, retain an independent reference stream and measure feedback drift without declaring retraining itself a benefit.

[PD] Transport requires explicit assumptions about workload, implementation fidelity, interfaces, authority and effect modifiers. Use target-population pilot data to test these bridges; a successful precursor elsewhere is not a transferable effect size. Bareinboim/Pearl provides the causal distinction, not an identified transport formula for this case.

Basis: proposed measurement [PD], bounded inference [AI]. Cohen (1960); Lakkaraju et al. (2017); Perdomo et al. (2020); Bareinboim/Pearl (2016).

TECHNICAL ANALYSIS

Evaluate the entire control path and its strongest alternative

[PD] Compare an approved capacity-aware control with current approved practice, a cheaper bounded check and a staffed alternative. Define feasible action rights and common evidence before selecting a policy.

[PD] Start with a shadow replay using controlled known states; then, only after authority and timing validation, randomize operating domains that contain shared queues, crews and spillovers. Retain all original route bookings, stopped checks and unresolved parcels. Existing urgent escalation and customer remedies continue in every arm. Do not introduce known dangerous conditions to create labels. Separate a physical workflow trial from the current synthetic decision analysis.

[PD] The physical estimand is $E[Y(\text{usual}) - Y(\text{control})]$ per original booking, where Y indicates nonreceipt by the original promise. The resource estimand is $E[L(\text{usual}) - L(\text{control})]$ per original route, where L prices all original failures and performed work; subtract additional fixed assigned cost K in the same route units. Report ordinary-flow harm on its original ordinary-unit denominator and overall harm on all originals. The replay reports failures per fixed ten-parcel route; divide by ten for booking rates. Detector accuracy, activity completion and changed membership are intermediate measures.

[PD] The compact prospective law varies observed dispatch slack and unobserved shared workload. Policies see slack and completed-check results, not future check durations, absence or recovery outcomes. A bounded policy starts only when the known maximum duration fits; ordinary lateness is structurally absent under that assumption, not empirically proven safe. Price every attempted check, failed recovery, reserved service and controller, even when no check starts. Compare no added gate, single and repeated checks, bounded variants and the best rule using the same permitted evidence.

[PD] Exact shared-domain moments and selected finite tails audit the declared synthetic law. Normal power remains a conditional approximation. Compare detection of an economic break-even effect with full-law clearance; large assumed effects cannot justify an underpowered material-effect test. Use independent pilot variance, timing, within-domain dependence and recruitment estimates before fixing a confirmatory design. Prespecify capacity, slack, parcel mix and runtime version as baseline modifiers; do not stratify total effects on realized holds or dispatch survival.

[PD] Costs include recurring maintenance, retraining, reserved capacity, false alerts, rework and displaced legitimate work. Operational credit, avoided resource use and released cash need separate bridges. A baseline rate lower bound or an ordinal effort label cannot establish a finite monetary entitlement or a return-on-cost ranking.

Basis: proposed estimands and design [PD]. Hudgens/Halloran (2008), Prentice (1989), Spearman et al. (1990). No source-calibrated probabilities or achieved effects.

TECHNICAL ANALYSIS

Specify when evidence rejects a claim or holds deployment

[PD] Falsifiable propositions distinguish a contradicted mechanism from a funding decision held because evidence is incomplete.

Proposition	Prespecified rejection or hold rule
P1: lower recorded defects imply better service	Reject if original-cohort deadline failures worsen beyond the declared margin despite improved recorded defects. The FIFO proof refutes the general implication.
P2: upstream origin bounds downstream recovery	Reject only when linked same-population downstream recovery exceeds the proposed numerical bound. One recovered unit refutes only zero interception.
P3: the gate preserves service capacity	Reject when observed check-time, false-hold and backlog paths breach prespecified original-deadline limits; model violations require a changed policy, not extra sample size.
P4: feedback improves the original outcome	Require a frozen-version comparison on an independent original-cohort reference stream; calibration on surviving labels alone is insufficient.
P5: the preferred control merits purchase	Require full-cost clearance and absolute harm limits against the strongest cheaper alternative. Imprecision is a hold, not zero effect.

[HR/RC] The private inventory contains 47 vectors: one weak contextual source pair, 16 version comparisons, 19 design trajectories and 11 profiles. The source pair has no fitted X or moving-range crossing and establishes no stable phase. One countdown design has a fitted X crossing, not an operating signal. Tentative XmR uses 2.66 and 3.267; no eligible Laney panel exists and none is computed. Missing weekly, field and retraining observations are not fabricated.

[PD] Under an assumed linear conditional-mean model, expected untreated follow-up after selection at baseline x is $m + \rho \times (x - m)$, with population mean m and baseline persistence ρ . This illustrates regression to the mean; it is not estimated here. A later observational rollout needs pretrend plots, equivalence bounds, negative controls and sensitivity analyses. Nonsignificant pretrends do not prove parallel trends.

Limitations and identifying assumptions

[AI/PD] Operating exposure, original-promise outcomes, funding and implementation are unauthenticated. Cause shares, sensor validity, transportability and additive prevention are unverified. Synthetic duration bounds, recovery and prices require calibration; violations can create ordinary-flow harm absent from the toy law. Queues, spillovers, selective reference and delayed labels remain material. Architectural completeness and arithmetic closure establish neither execution nor causal effectiveness.

Basis: source/SPC boundaries [HR/RC], exact new counterexample [AI], proposed falsification and evaluation [PD]. Barnett et al. (2005); Lakkaraju et al. (2017).

PROPOSED OPERATING TEST / 1 OF 4

Spend time only when the original promise can absorb it

[PD] This finite replay compares no added check with up to two physical checks of the same parcel. It is not a live gate or a historically calibrated forecast.

Primitive	Base / tight-slack	Staffed redesign
Original route	1 critical + 9 ordinary parcels	Same original 10
Shared hidden grade	0.60 / 0.40	Same
Critical parcel absent by grade	0.20 / 0.40	Same
Check duration; long probability	1 or 3; 0.20 / 0.60	1 or 2; same probabilities
Completed confirmation/recovery	0.75 / 0.55	0.90 / 0.80
Original slack 2 / 4	0.50 / 0.50; stress 0.80 / 0.20	0.50 / 0.50
Work price; two-check fee	1 per time unit; 0.15	1.5 per time unit; 0.60

[PD] Three routes share one hidden grade; domains are independent. Within grade, routes and primitive absence, durations, completion draws and slack are independent. Policy observes own-route slack and completed results, never absence or future duration. A check starts only if its known maximum fits remaining slack; stop on successful physical confirmation/recovery. The second attempt addresses the same unresolved parcel, not a second credited benefit.

[PD] Without delay, only an absent unresolved critical parcel fails. A dispatch overrun makes all ten miss their original promise. Price every original failure at 12 units and all performed check time, including present parcels and failed checks. Main fees are paid on every route, including unused capacity.

THE DURATION BOUND IS A TESTABLE ASSUMPTION

[PD] The bounded law has zero ordinary lateness by construction. It does not demonstrate field safety. Unbounded checks are shadow stress comparators. Unknown authority or duration requires the approved fallback; the toy supplies no permission to hold customers.

[PD] Wholly invented. Original promise +28 days independently reconstructs receipt timing; later recovery does not erase a miss. Common existing controls/remedies stay. Reference study cost is separately budgeted.

A better mean does not establish enough economic precision

[PD] Gates require a reduction in variable burden exceeding fixed fee K and fewer original-deadline failures. All original routes and their parcels remain in both endpoints.

Domains / arm	80% MDE	Fee K	Zero-test power at BE	Full-law joint floor
Small / 16	not attained	0.15	2.8%	0.0%
Base / 128	not attained	0.15	4.8%	0.0%
Tight slack / 128	not attained	0.15	unreachable	0.0%
Staffed / 256	0.7570	0.60	55.9%	8.4%

[PD] The assumed response path uses two bounded checks with probability λ per original route, no added check otherwise; K stays fully paid. Exact shared-grade moments set fixed normal radii at nominal directional 2.5%, using the largest variance over λ in $[0,1]$. These are conditional planning screens, not achieved field power or uniform error guarantees.

[PD] Base maximum variable gain 0.24784 exceeds fee 0.15 but is too small for the selected precision. Tight-slack maximum gain 0.099136 is below that fee: no sample can create break-even on this path. Staffing changes durations, completion, price and fixed capacity; its MDE 0.75704 exceeds fee 0.60. Break-even detection is only 55.9%.

[PD] Eighteen exact marginal tails retain discreteness. Base loss/service-null errors are 2.50691%; staffed 2.37247%. Economic-boundary false clearance is 2.05291%/1.03605%. At full law, exact economic clearance is 3.23965% base and 8.51578% staffed; their joint lower bounds are 0% and 8.51578%. Zero is an uninformative lower bound, not zero joint power.

DO NOT BUY PRECISION FOR THE WRONG COMPARATOR

[PD] The base common-slack rule has 0.08516 lower total cost than the main controller. The staffed controller equals its common-slack optimum, yet its 1,536-route / 15,360-booking design still fails economic precision. No allocation is recommended as confirmatory or operationally feasible.

[PD] Four direct allocations, no minimum search or Monte Carlo. Joint bounds use the union bound on exact marginal tails, not an exact joint distribution. An approved ordinary-flow harm threshold and duration-overrun model are still required.

PROPOSED OPERATING TEST / 3 OF 4

Pay for the entire check and preserve the cheaper alternative

[PD] Entries are total resource cost including fees / original failures per ten-parcel route. These are invented resource units, not dollars or observed customer rates.

Policy	Base	Tight slack	Staffed
No added check	3.3600 / 0.2800	3.3600 / 0.2800	3.3600 / 0.2800
One unbounded	24.2988 / 1.8774	37.0817 / 2.9426	3.1680 / 0.0440
Two blanket	32.4835 / 2.5090	45.2493 / 3.5728	8.3244 / 0.4487
One bounded	3.2020 / 0.1910	3.3268 / 0.2444	3.1680 / 0.0440
Two bounded	3.2622 / 0.1741	3.4109 / 0.2376	3.0996 / 0.0258
Best own-slack rule	3.1770 / 0.1910	3.2868 / 0.2444	3.0996 / 0.0258
Perfect grade / bounded	3.1770 / 0.1910	3.2868 / 0.2444	2.9664 / 0.1392

[PD] One-check and blanket fees are 0.05 base/stress; every staffed checking plan pays 0.60, even if unused. Own-slack optimization enumerates all nine mappings over no/one/two bounded plans before capacity reservation; it pays the chosen plan fee. Other routes cannot reveal grade. Blanket rules are diagnostic, not authorized customer policies.

[PD] Base net versus no check = $a \cdot h \cdot 1.27104 - 1.02320 - 0.15$. Favorable avoided failures receive realization a and nonduplicated share h ; every check, added failure and fee stays fully charged. Same-parcel benefits count once under the declared common-primitive coupling.

Net units / route	$h = 0.60$	$h = 0.80$	$h = 1.00$
$a = 0.50$	-0.7919	-0.6648	-0.5377
$a = 0.75$	-0.6012	-0.4106	-0.2199
$a = 1.00$	-0.4106	-0.1564	0.0978

[PD] At failure price 6, base one/two cost 2.0560/2.21768 versus no check 1.6800; neither pays. At price 24, one/two cost 5.4940/5.35112 versus no check 6.7200. The preferred action depends on a prospectively agreed price, not an estimated historical dollar saving.

[PD] Primary bounded rules add no deadline harm within this law; blanket checks do, and charge it in full. Reference reassignment within route-burden support 0..18 on epsilon original routes per arm permits $36 \cdot \epsilon$ burden error; with verified work fixed, $24 \cdot \epsilon$. Outcome error permits $2 \cdot \epsilon$ critical-failure contrast error.

PROPOSED OPERATING TEST / 4 OF 4

Use the information ceiling before buying another diagnostic

[PD] With the base prices and bounded menu, knowing hidden workload grade cannot improve the resource decision: use one check when slack is four, otherwise no added check. Its information value is zero.

[PD] Grade information leaves future absence, check duration and completion unknown. Under staffing it can reduce expected total cost from 3.0996 to 2.9664, a ceiling of 0.1332 before any information fee. But failures rise from 0.0258 to 0.1392 per route because the cost-optimal rule sometimes declines checks. This is not a service-constrained optimum or a permission to accept the trade-off.

[PD] A separate research choice has states: duration/spillover validation useful (0.40), recovery validation useful (0.35), neither (0.25). Payoffs for those studies and deferral are (5,-2,-1), (-2,7,-2), (0,0,0), including downstream study consequences. Signal accuracy q has symmetric errors; enumerate all 27 rules.

Research scenario	q	EVSI	Cost	Net EVSI
Low	50.0%	0.71250	0.40000	0.31250
Base	75.0%	1.90625	1.00000	0.90625
High	90.0%	2.74250	2.00000	0.74250

[PD] No-information value is 1.15 and EVPI 3.30. These are alternative hypothetical likelihoods and prices, not confidence limits. The research objective differs from buying grade information for the existing controller. Its utility is not added to operating savings or to the controller comparison.

VALIDATE BEFORE AN AUTHORIZED FIELD TEST

[PD] If only a fraction rho of original routes changes after an unexpected overrun, added failure burden is at most 120*rho, plus all extra work. At rho=0.001, the 0.12 allowance exceeds base net 0.09784. This is a conditional worst-case sensitivity, not an event probability; wider spillover invalidates that frame. Audit duration tails, false confirmations and missing receipt, and agree a service constraint before testing.

[PD] The grade-information ceiling applies only to its frozen menu, prices and predecision rights. All-cause field harm may exceed the toy support; the ordinary-parcel denominator and stratum/linkage uncertainty require separate bounds.

REPRODUCIBILITY / READER PROTOCOL

Reproduce the physical flow and the original service promise

[AI] The Flow-and-Service Evidence Contract tests whether a control reduces a real failure without concealing the customers delayed or excluded by it.

Reproduction layer	Required record and falsification test
Original cohort and clock	Freeze eligibility and promised completion before assignment. Retain held, rerouted and removed items; distinguish finished late from unfinished at the observation boundary. A changed itinerary cannot erase the original promise.
Physical flow and capacity	Record arrival, service, interruption, queue discipline, backlog and disposal. Reconcile every original item. Test control time, false holds, rework and displaced ordinary work against available capacity.
Control and dependencies	Version the attempted check, execution dose, observed evidence, failure response and release authority. Separate named stage coverage from tested effectiveness. Include prerequisites and a safe, authorized fallback.
Net customer and resource effect	Compare the proposed configuration with feasible continuation on mature original outcomes. Separate recorded defect removal, physical repair, deadline attainment, avoidable expense and cash release.
Learning and decision	Preserve all outcome classes in the learning sample, including control-induced delays and unresolved cases. Lock the model and decision rule before testing; keep counterfactual truth distinct from information available at action time.

[PD] Reproduce baseline and control traces with identical original arrivals and deadlines. Then change capacity, inspection duration, recovery likelihood and available alternatives. Compare a full bundle with the smallest coherent intervention and separately priced capacity. Missing prerequisite evidence requires a decision hold; it does not establish that every partial intervention is worthless.

[AI] The private archive contains source versions, recovery boundaries, calculation code, decision assumptions and internal review. Public examples are independent constructions. They neither reveal historical operating coordinates nor turn a proposed control into evidence of deployment.

Protected registry and bounded recovery record; proposed FSEC reproduction protocol. Identities, internal quantities and source locations remain private.

REFERENCES / 1

Primary sources and their role in this case

[AI] These sources inform flow control, measurement, learning and evaluation. They do not certify the proposed reduction, funding range or operating result.

NIST/SEMATECH. Individuals Control Charts, e-Handbook section 6.3.2.2. Comparable operating observations are required for process limits. A proposed wave schedule, sensitivity vector or alternative baseline is not a stable time series.

Laney, D. B. (2002). Improved Control Charts for Attributes. *Quality Engineering*, 14(4), 531–537. Use authentic defect counts and exposure for rational subgroups. A missing repeated-count panel cannot be reconstructed from rounded rates or funding scenarios.

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20, 37–46. Independently code action delivery, physical custody, original deadline and final customer status. Agreement is not proof that a workflow state represents the intended outcome.

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. Worst-unit selection can concentrate temporary extremes. Freeze selection and compare untreated units before attributing regression toward typical performance to a new gate.

Callaway, B., & Sant’Anna, P. H. C. (2021). Difference-in-Differences with Multiple Time Periods. *Journal of Econometrics*, 225, 200–230. In a later phased deployment, estimate group-time effects on original customer cohorts with coherent baseline and maturity. Shared queues, treatment versions and anticipatory changes require explicit assumptions.

Collier, D. (2011). Understanding Process Tracing. *PS: Political Science & Politics*, 44, 823–830. Attribute architecture and analysis separately from funding, staffing, installation and operational results. A record of proposed ownership does not establish accepted delivery responsibility.

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. Price information by the feasible control configuration it changes. Separate decision-study value from hypothetical customer benefits and include cheaper bounded checks.

Primary author, journal, institutional and official methods sources. Read scope and case-specific application are retained privately.

REFERENCES / 2

Primary sources and their role in this case

[AI] These sources inform flow control, measurement, learning and evaluation. They do not certify the proposed reduction, funding range or operating result.

The W. Edwards Deming Institute. The Funnel Experiment. Adjusting operations in response to routine variation can create extra variation. Establish measurement and process phases before changing a gate or throttle.

Campbell, D. T. (1976). Assessing the Impact of Planned Social Change. Rewarding disappearance of a defect label can shift recording or eligibility. Keep held and removed items in the original customer cohort and observe final service independently.

Goodhart, C. A. E. (1992). The Objectives for, and Conduct of, Monetary Policy in the 1990s. Reserve Bank of Australia Annual Conference, section 1(b)(i). A target can change the link between a measure and its intended result. A lower dispatched-item defect rate can coexist with more delayed or excluded customer orders.

Ittner, C. D., & Larcker, D. F. (2003). Coming Up Short on Nonfinancial Performance Measurement. *Harvard Business Review*, 81(11), 88–95, 139. Test the link from measured compliance through delivered control to mature service and avoidable resources. A stage map and an arithmetically reconciled plan cannot validate that chain.

Tucker, A. L., & Edmondson, A. C. (2003). Why Hospitals Don't Learn from Failures: Organizational and Psychological Dynamics That Inhibit System Change. *California Management Review*, 45(2), 55–72. Local recovery workarounds can conceal design problems. Preserve frontline exceptions and feedback, with authority to repair the underlying process rather than merely close tasks.

OECD / European Commission JRC (2008). Handbook on Constructing Composite Indicators: Methodology and User Guide. Coverage weights and aggregation rules are modeling choices. An overlap haircut does not estimate interactions, shared causal effects, capacity loss or implementation feasibility.

Spearman, M. L., Woodruff, D. L., and Hopp, W. J. (1990). CONWIP: A pull alternative to kanban. *International Journal of Production Research*, 28(5), 879-894. doi:10.1080/00207549008942761. Evaluate released work, unreleased backlog, capacity and promised service jointly; controlling WIP does not create free capacity.

Primary author, journal, institutional and official methods sources. Read scope and case-specific application are retained privately.

REFERENCES / 3

Primary sources and their role in this case

[AI] These sources inform flow control, measurement, learning and evaluation. They do not certify the proposed reduction, funding range or operating result.

Prentice, R. L. (1989). Surrogate endpoints in clinical trials: Definition and operational criteria. *Statistics in Medicine*, 8(4), 431-440. doi:10.1002/sim.4780080407. A change in a leading control signal is not sufficient evidence of improved original-promise receipt when other harm paths remain.

Bareinboim, E., and Pearl, J. (2016). Causal inference and the data-fusion problem. *Proceedings of the National Academy of Sciences*, 113(27), 7345-7352. doi:10.1073/pnas.1510507113. Specify which population and implementation mechanisms differ before transporting evidence from another workflow or network.

Lakkaraju, H., Kleinberg, J., Leskovec, J., Ludwig, J., and Mullainathan, S. (2017). The Selective Labels Problem: Evaluating Algorithmic Predictions in the Presence of Unobservables. *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 275-284. doi:10.1145/3097983.3098066. Retain held, removed and dispatched original units in outcome reference sampling; a changed decision rule changes which outcomes become training labels.

Perdomo, J., Zrnic, T., Mendler-Dunner, C., and Hardt, M. (2020). Performative Prediction. *Proceedings of the 37th International Conference on Machine Learning*, PMLR 119, 7599-7609. Freeze feedback-model versions and measure independent original-cohort outcomes; retraining on policy-altered observations does not itself establish improvement.

Zimmerman, T. A. (2019). Manufacturing Profile Implementation Methodology for a Robotic Workcell. NISTIR 8227, sections 4-5 and 7. Compare individual configurations and the completed control bundle against a characterized baseline; preserve configuration and event records. The robotic cybersecurity testbed is a methodological analogy, not evidence that these proposed delivery controls work.

Hudgens, M. G., & Halloran, M. E. (2008). Toward Causal Inference with Interference. *Journal of the American Statistical Association*, 103, 832-842. Original manuscript. Assign domains containing shared queues and capacity spillovers, or explicitly estimate interference.

Primary author, journal, institutional and official methods sources. Read scope and case-specific application are retained privately.

FINAL REFEREE REPORT / INTERNAL REVIEW

Make controls earn their place

[AI] Author-led, agent-assisted review; not external peer review. The case challenges the transition from a stage map to an executable improvement claim.

Major issue	Resolution and decision boundary
Stage coverage treated as causal completeness	Map failure creation, detection and interception separately. A downstream check may catch upstream failures; one assigned programme per stage does not prove completeness or effectiveness.
An overlap haircut treated as a joint model	Keep scenario arithmetic explicit. Dependency, substitution and interference require a response law and evidence. A tested sensitivity point is not a guaranteed lower bound.
Recorded prevention treated as customer recovery	Retain original deadlines and cohorts. The finite replay distinguishes final backlog from all deadline failures; completed-only reporting can hide both waiting and late completions.
A hard gate treated as costless assurance	Count service time, false holds, displaced work and control failures. Compare bounded checks and priced capacity. Enforcement does not make imperfect detection or execution infallible.
A funding range treated as a benefit bridge	Separate cost from benefit, preserve scope and align the horizon. Imported baseline scenarios cannot replace absent plan inputs; a lower-bound baseline supplies no finite upper-value bound.
A feedback loop treated as guaranteed learning	Observe harms outside the target defect label, version changed policies and retain unresolved outcomes. A loop trained on its own selected labels can miss the losses it creates.

Partner critique: can a COO act on page one?

[AI] Yes. Name the original service obligation, the permitted control and its capacity budget. Test the smallest coherent configuration against feasible alternatives, with clear stop rules for customer delay and incremental resources. Preserve useful architecture while requiring a separate valuation and delivery owner.

Remaining limits and minor issues

[AI] The technical layer consolidates source authentication, measurement, transport, implementation, economics and reference limits. Neither the finite queue nor the workflow replay estimates historical performance. Internal checks support a bounded design case; they do not establish accepted investment or delivered benefit.

REVIEW DISPOSITION / PASS WITHIN STATED SCOPE

[AI] Publish the evidence reconciliation, reproducible counterexamples and prospective test. Advance a control only through the appropriate operating authority with measured capacity and mature customer outcomes.

Final source, independent numerical, methods, partner and actual-page reviews required. Internal publication review is separate from external assurance.