

OPERATING CASE STUDY / CUSTOMER EXPERIENCE & LOGISTICS

Workflow Analytics & Process Validation

Make a rapid-completion flag useful without overstating what it proves

An anonymized analytical case examining a short-duration workflow detector, its audit priorities and its unsupported cost-leak hypothesis. It reconciles event populations and designs a test of verified process improvement.

PROJECT CONTEXT / PROCESS INTEGRITY & AUDIT DECISION DESIGN

BLUF / BOTTOM LINE UP FRONT

Validate the signal before routing action.
Treat rapid completion as an audit lead.

ANALYTICAL FINDING

Under the record's negative-error interpretation, the counts cannot support a clean filtered population. Recover error definitions and reconcile cohorts before authorizing action.

CONTRIBUTION IN THE SOURCE RECORD

Designed the diagnostic and audit ranking; retained contrary outcome evidence and withdrew the cost claim. Provisional registry hoop support.

EVIDENCE STATUS

Originals unrecovered. Detector accuracy, corrective implementation and realized benefit remain unestablished.

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

EXECUTIVE ABSTRACT / DECISION AND OWNERSHIP

Commission a validation gate before expanding signal-led audits

[AI/PD] The proposed Workflow Verification Contract (WVC) separates an event-time flag, an independently verified protocol defect, an authorized corrective action and a later operating outcome. Each step needs its own population and evidence.

Abstract element	Decision-relevant synthesis
Context [HR]	A short-duration workflow analysis replaced anecdotal concern with a station-level audit list. The author examined hidden filters, joined external metrics, revised a composite ranking and traced downstream records.
Objective [PD]	Determine whether the detector can direct limited audit capacity toward verified, preventable process defects. Avoid equating fast completion, administrative labels or a high rank with misconduct or financial loss.
Approach [RC/AI]	Reconcile event counts and exclusions, test timestamp and entity validity, distinguish post-action selection from confounder control, assess ranking sensitivity and specify a prospective audit-policy comparison.
Findings [AI]	The negative-error interpretation fails a necessary count bound for aligned populations. Outcome-conditioned comparisons remain selected. Lower recorded cost and fewer observed repeats establish neither savings nor absence of unresolved defects.
Decision value [PD]	Authorize a bounded data and reference-audit phase. Release a ranking for operational use only after its event basis and predictive validity are verified; compare prioritization against an equal-capacity alternative.
Implications [PD]	Data owners certify event semantics and exclusions. Quality owns independent adjudication and corrective authority. Operations owns audit capacity; an evaluator measures later defects, customer outcomes and full cost.

EXECUTIVE ACTION AND THRESHOLD

[PD] Hold unvalidated corrective-action assignments. Require reconciled populations, valid timestamps, a frozen scoring rule and independent labels from flagged and unflagged records. Scale an audit policy only after it improves verified outcomes and clears net-value and safety gates.

Basis: S1 selected project register. [HR] historical record; [RC] recalculated; [AI] analytical inference; [PD] proposed design. A screening flag does not establish intent.

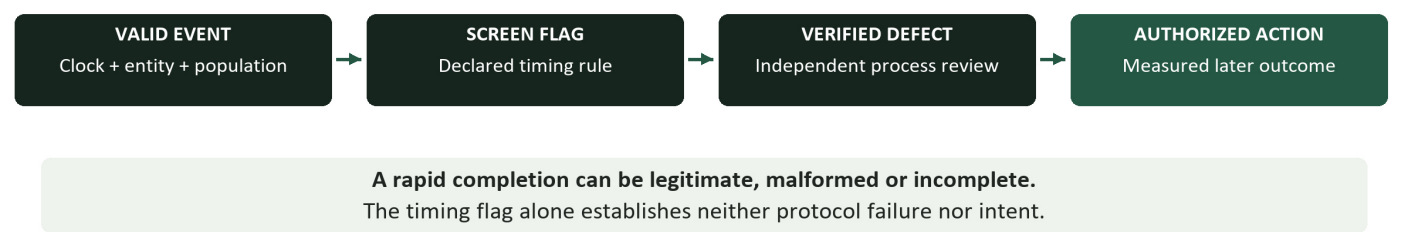
EXHIBIT 1 / SIGNAL-TO-ACTION CONTRACT

Separate fast completion from the defect an audit is meant to find

[HR] The source created a short-duration flag after an earlier misuse report lost validity. It combined flag rate, flag volume and other site metrics to produce an audit list. [AI] The flag identifies a timing pattern; its meaning depends on the workflow, event clock and independent reference evidence.

A VALIDATED SIGNAL HAS FOUR SEPARATE STATES

Proposed Workflow Verification Contract / progression requires evidence and accountable ownership



Basis: S1 detector and contribution record; proposed state contract. No individual misconduct inference.

State	Evidence needed to advance
Valid event [PD]	A unique event with documented opening, resolution and capture timestamps, unit conventions and exclusions. A fast elapsed value can be legitimate, malformed or incomplete.
Screening flag [HR/PD]	A fixed rule applied to the declared eligible population. Retain missing and invalid records separately; do not silently convert them into fast or normal events.
Verified protocol defect [PD]	Blinded review against a versioned process standard and independently retrieved evidence. Record disagreement and uncertainty; speed alone does not establish bypass or intent.
Authorized correction [PD]	A responsible process owner selects a justified action. Evaluate its consequences using mature outcomes on original eligible cohorts, including records not selected by the screen.

A SCREEN IS AN INVESTIGATION AID

[AI/PD] Use aggregate process and site signals to allocate diagnostic attention. A verified process defect still does not establish an individual’s intent. The portfolio case supports a validation design, not an automated personnel decision.

Basis: S1 detector definition, ranking method and ownership boundaries. Proposed WVC states distinguish measurement, diagnostic validation and action.

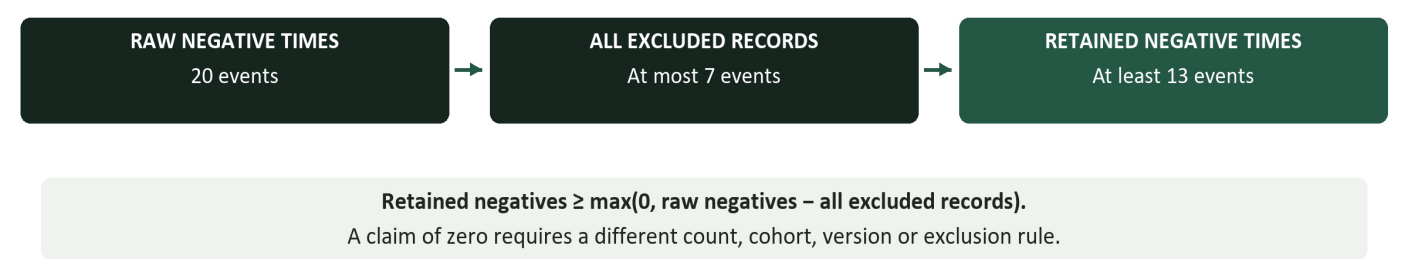
EXHIBIT 2 / POPULATION CONSERVATION

Check whether the reported exclusions can remove the claimed error count

[RC/AI] The register interprets an error category as negative durations and claims none remain after exclusion. Under that interpretation and one aligned event universe, the counts cannot all hold. Recover the error meaning and population bridge before accepting a clean signal.

AN EXCLUSION CANNOT REMOVE MORE ERRORS THAN RECORDS

Wholly hypothetical count example / a necessary nesting condition, not a fitted project estimate



Basis: original conservation-bound analysis. Historical counts and source-version forks are private.

[AI] Let N be negative-duration records in a raw window, X all records excluded from that window, and R retained negative-duration records. If raw and retained populations are nested under the stated exclusion, $R \geq \max(0, N - X)$. Excluding X records cannot remove more than X negative records.

Interpretation	Decision consequence
Common cohort and version [AI]	If the error category truly means negative durations and counts share an event universe, the zero-retained-error claim fails. Recompute the count and downstream ranking effects.
Different cohort or version [AI]	A later cleanup, changed timestamp field or additional exclusion could explain the discrepancy. Recover that transformation and its count bridge before combining the results.
Different error meaning [AI]	The detailed error category may include nonnegative mismatches. Recover its row-level definition before applying the negative-count bound; retain this alternative alongside cohort and version changes.

WHAT THE FINDING CHANGES

[AI] High overall label agreement does not settle error meaning or tail validity. The next investment is a reproducible event-universe reconciliation, before coaching, ranking deployment or a benefit commitment.

Basis: S1 M02/M06/M16/M17 and version reconciliation. Public diagram is wholly hypothetical; confidential source counts and the derived lower bound remain private.

EXHIBIT 3 / EVENT CLOCKS AND MISSING RECORDS

Validate the timing tail on the population that receives the decision

[HR] Alternative timestamp interpretations produced different window mixes before the source preferred the workbook’s own label. It also documented negative durations, invalid labels and a unit-name inconsistency. [AI] A high aggregate match rate can coexist with a material error in the specific tail driving audits.

VALIDATE THE TAIL THAT DRIVES THE AUDIT

Proposed event-quality contract / aggregate agreement cannot certify a decision-critical subgroup



Retain invalid and missing records as explicit accounting states.
Check flagged and unflagged events on the same independent reference.

Basis: S1 M16/M17; proposed reference audit. A field label is not a verified unit definition.

Validation question	Required evidence
What interval is measured? [PD]	Identify physical task start and completion, workflow opening, scan capture and posting times. Verify time zone, clock synchronization and the field’s actual unit, rather than trusting its label.
Which events are missing? [PD]	Retain invalid windows, negative durations and unknown dispositions in the accounting bridge. Compare error prevalence across process types and sites before excluding records from a ranking.
Does the cutoff behave as intended? [PD]	Audit exact-boundary, zero-duration, duplicate, delayed-posting and reopened events. Predefine interval endpoints and whether one physical task can generate several workflow records.
Is the joined window aligned? [HR/PD]	The event extract covers a subset of the reporting cycle while joined site metrics use a broader window. Align dates and eligible exposure or preserve the comparison as contextual only.

TAIL VALIDITY IS A SEPARATE GATE

[PD] Sample flagged and unflagged records across sites and process types, oversampling known error modes with analysis weights. Report tail-specific predictive accuracy and missing-label bounds; do not substitute overall timestamp agreement for defect validation.

Basis: S1 M16/M17 and extraction-integrity record. Event-log quality and verification methods guide the proposed reference audit.

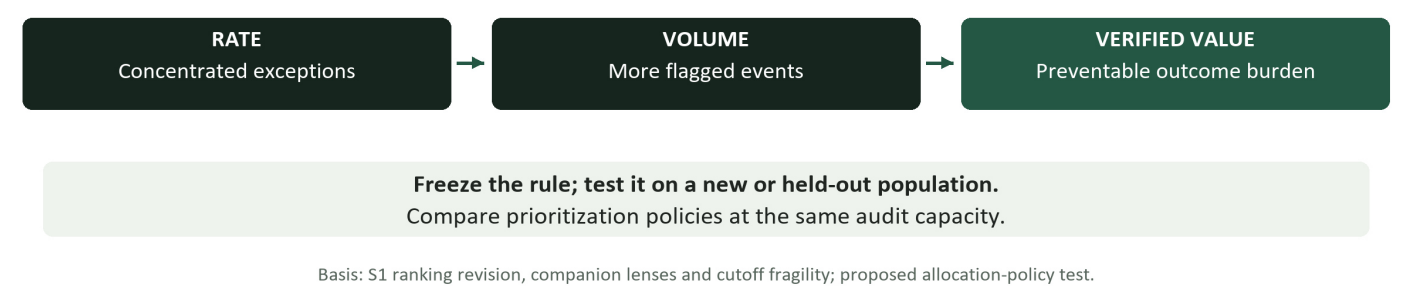
EXHIBIT 4 / RANKING AND DECISION CAPACITY

Treat the composite ranking as a capacity-allocation policy

[HR] A verifier challenge led the author to raise the volume component after a small site displaced higher-exposure sites. Separate rate and volume lists were retained. [AI] That is a choice about audit priorities; it does not reveal an objectively correct ordering or calibrate a probability of misuse.

CHOOSE THE AUDIT OBJECTIVE BEFORE TUNING THE SCORE

Historical score revision + proposed policy validation / no confidential weights or site names shown



Ranking choice	Test before operational release
Rate versus volume [HR/AI]	High rate and high count identify different audit objectives. State whether the policy seeks concentrated protocol failures, the largest preventable burden or a representative diagnostic sample.
Scaling and weighting [PD]	Freeze the eligible population, transformation, reference distribution, weights and clipping rule. A site’s standardized score can change when other sites enter the reference set.
Cutoff uncertainty [HR/PD]	The last selected and first excluded scores are close in the source. Report a near-cutoff review band and sensitivity to reasonable weights, timing corrections and missing records.
Validation after tuning [PD]	Evaluate the frozen rule on a later or independently held-out population. A weight changed after inspecting an inconvenient result cannot count as an independent validation on the same data.

COMPARE POLICIES AT THE SAME AUDIT BUDGET

[PD] Test signal-prioritized audits against an equal-capacity alternative. Count confirmed defects, unnecessary reviews, corrective workload and later verified outcomes. More flags or a more persuasive top-site list is not the objective.

Basis: S1 composite-score revision, companion lenses and rank-boundary warning. Ranking weights and the exact internal cutoff remain private.

EXHIBIT 5 / SELECTION AND OUTCOME INTERPRETATION

Within-bucket comparisons can remain selected after process adjustment

[HR] The analysis moved from raw fast-versus-slow comparisons to process-by-outcome cells and partial correlations using bookings or disposal. [AI] Process adjustment can address one compositional difference. Conditioning on an outcome or downstream disposition can also create a new selection problem.

A SELECTED SUBGROUP CAN REVERSE THE OBSERVED COMPARISON

Synthetic collider example / independent Bernoulli(0.5) F and C; booking B = F OR C

TRUE INTERVENTION EFFECT
Y = C + 0.20F: effect +0.20

≠

WITHIN BOOKED EVENTS
Mean Y: 1.00 → 0.70

F → B ← C: selecting booked events changes the severity mix.
Apparent effect -0.30 can coexist with a true effect +0.20.

All numbers are hypothetical. This proves a possibility; it does not identify the historical mechanism.

Comparison	What remains unresolved
Raw fast versus slow [HR/AI]	The groups contain different process mixes and exception severity. An aggregate association cannot isolate what faster resolution changes.
Within a booked category [AI]	The category may depend on latent defect severity, action, observation and recognition. Restricting to it can open a noncausal path; the comparison is not automatically selection-free.
Given disposal or bookings [AI]	These variables can be mediators or colliders. Adjustment may change the estimand or induce bias. A larger partial correlation is not proof of a hidden causal leak.
Weighted rate product [HR/AI]	Multiplying a site outcome rate by an event-level flag share combines different denominators. It can be a declared prioritization index; it is not an attributable defect count or avoidable-cost estimate.

PRESERVE THE CONTRARY EVIDENCE

[HR/AI] The author withdrew the cost-leak claim when the observed dollar pattern did not support it. Retain that decision. The opposing pattern also does not establish savings: treatment assignment, comparable severity and complete customer truth are missing.

Basis: S1 M04/M07–M12/M14 and challenge response. Selection-bias analysis applies to the proposed interpretation, not an estimated historical causal effect.

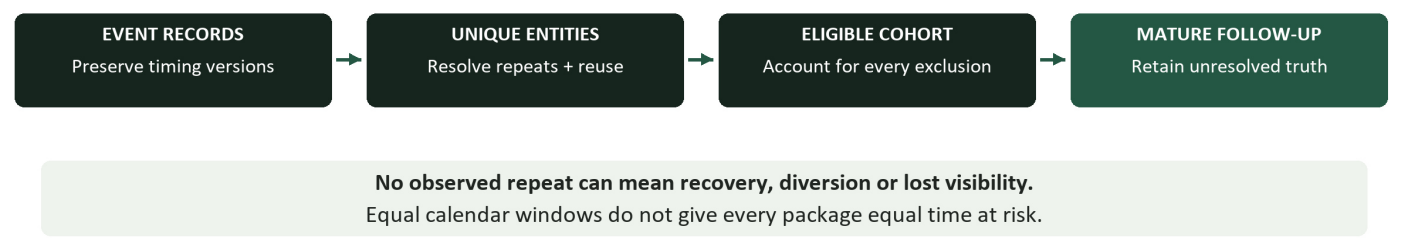
EXHIBIT 6 / ENTITY AND FOLLOW-UP

Follow original packages long enough to distinguish recovery from lost visibility

[HR] The source traced repeated package identifiers within the extract and compared first-touch groups. [RC/AI] The multiplicity histogram reconciles the event total, but the reported first-touch comparison includes fewer identifiers than the unique-ID inventory. Its scope and observation time must be explicit.

EVENTS, PACKAGES AND FOLLOWED COHORTS ARE DIFFERENT UNITS

Proposed entity and time contract / each transition needs a complete count bridge



Basis: S1 recurrence histogram and first-touch comparison. No survival or safety effect is estimated.

Reconciliation layer	Required rule
Events to entities [PD]	Link repeated records to the correct physical package and workflow episode. Preserve identifier reuse, merges, corrections and unmatched records; an event count is not a package count.
Entities to comparison [RC/PD]	Account for every unique identifier entering or leaving the first-touch comparison. Document invalid windows, exclusions and cohort rules before describing the comparison as exhaustive.
Time at risk [AI/PD]	A package first observed late in an extract has less time to recur. Start follow-up at the defined original event and use a common horizon or a justified censoring model.
Observed recurrence [AI/PD]	No later flag may mean resolution, diversion, lost visibility or insufficient follow-up. Ascertain customer condition and final disposition independently of the detector under evaluation.

DO NOT PROMOTE A SHORT-WINDOW PROXY INTO SAFETY EVIDENCE

[AI/PD] Lower observed recurrence does not prove successful recovery. Retain unresolved cases, measure competing dispositions and validate follow-up. Reopening an event and suffering a customer defect are separate outcomes.

Basis: S1 M13/M17 and event/identifier count reconstruction. Kaplan–Meier informs the censoring distinction; the incomplete source does not support a fitted survival model.

EXHIBIT 7 / CONTRIBUTION AND DECISION RELEASE

Credit the analytical detector without claiming ownership of the fix

[HR] The register attributes pivot-filter investigation, signal definition, score revision, verification and the challenge response to the case author. Separate owners were assigned reporting, network quantification and adoption work. The record does not establish implementation of the recommended fixes.

ANALYTICAL OWNERSHIP DOES NOT ESTABLISH FIX OWNERSHIP

Historical role boundary + proposed release responsibility



The record supports the diagnostic and the decision to withhold a cost claim.
Implementation of the recommended fixes and realized benefit remain unestablished.

Basis: S1 attribution and separate operating action owners. Provisional registry hoop support.

Claim or release gate	Evidence grade and decision owner
Analysis and ranking [HR/AI]	Provisional registry hoop support for authorship and the documented method. Authenticate the original workbook and scripts before strengthening the grade.
Contrary evidence and restraint [HR/AI]	Provisional registry hoop support for retaining the adverse cost finding and withholding a financial entitlement. This demonstrates a recorded decision boundary, not a realized saving.
Reporting and corrective delivery [HR/AI]	Assigned to other operating owners; completion remains unestablished. A proposed second-scan control, station audit or layout standard is not an implemented intervention.
Release authority [PD]	Data and Quality owners approve population validity and reference accuracy. Operations accepts audit capacity and corrective workload. An evaluator tests later outcomes and incremental value.

A DEFENSIBLE PORTFOLIO CLAIM

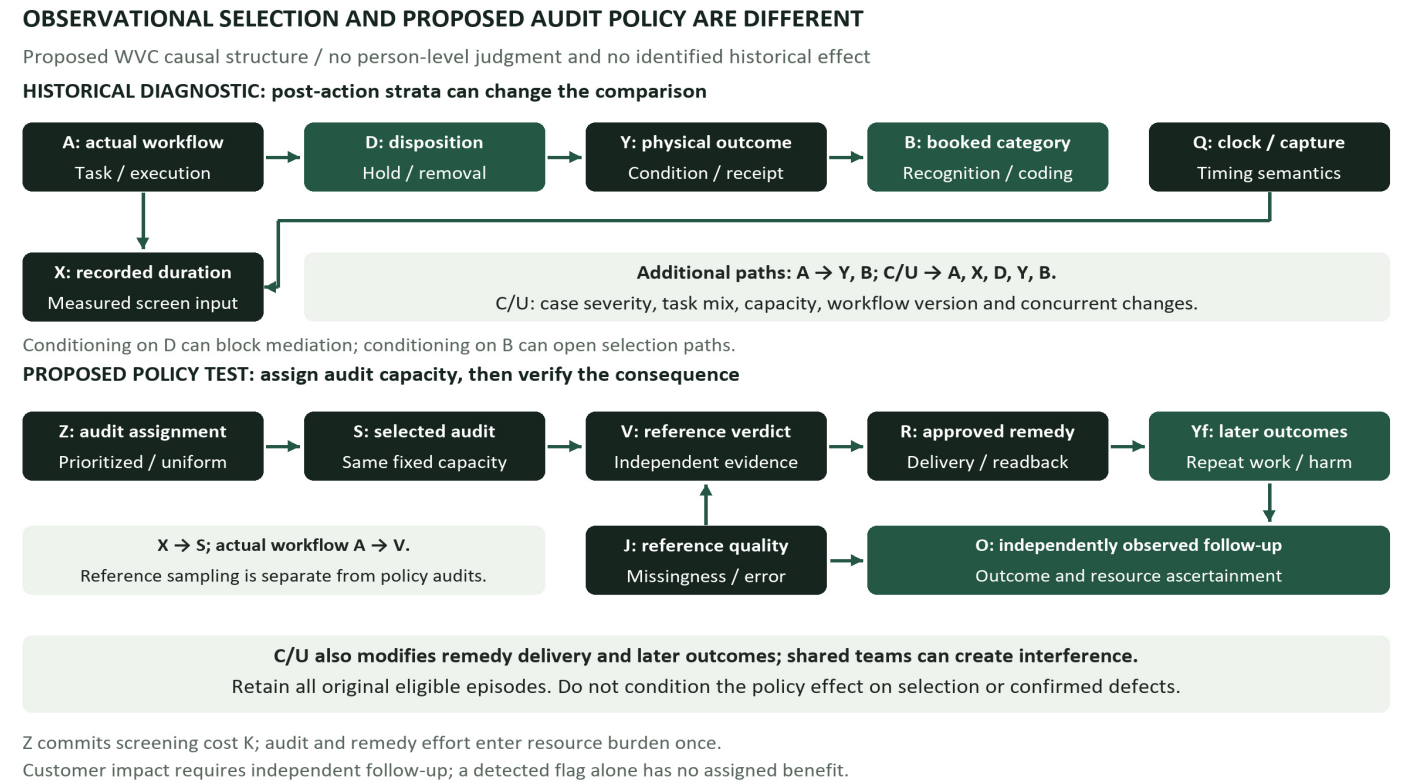
[AI] The supported capability is building and challenging an operational diagnostic, preserving contrary evidence and defining what must be verified next. The case does not label a person or site as having committed misconduct, or claim a cost leak that the source did not establish.

Basis: S1 personal contribution, action ownership and implementation limits. The challenger’s precise identity is unresolved and is not inferred from an inconsistent role label.

TECHNICAL ANALYSIS

Separate the observational screen from the policy that uses it

[AI/PD] The proposed Workflow Verification Contract (WVC) treats a timing flag as a candidate for review. It links validated measurement to independent evidence, authorized process correction and measured consequences; none of those transitions is established by a short duration alone.



[AI] In the historical panel, A is actual workflow execution; X is recorded duration; Q is clock and capture quality. D is disposition, Y physical outcome and B booked category. C/U includes severity, task mix, capacity, workflow version and concurrent changes. Restricting D or B can change the population or open a noncausal path.

[PD] In the policy panel, Z assigns audit allocation, S selects an audit, V is an independent verdict and R an authorized remedy. Yf is the later outcome; O is its independent observation through reference quality J. Z commits screening cost K. Audit and corrective effort enter the resource account once.

[PD] Estimate the policy effect across all original eligible episodes. Do not condition on a selected audit, a confirmed defect or completed correction. Use a separate reference sample for detector validation. The graph contains no person-level inference: a procedural deviation does not establish intent.

Basis: registered diagnostic paths [HR]; proposed WVC structure [AI/PD]. All arrows are hypotheses. Shared staff and queue capacity require an assignment boundary that contains spillovers.

TECHNICAL ANALYSIS

Reconcile the fastest tail before certifying the whole signal

[HR/AI] The source investigated a hidden filter and alternative clock interpretations. That is useful engineering work. A high overall agreement rate does not resolve incompatible retained-error claims or validate the physical inspection interval.

Gate	Required reconciliation [PD]
Event universe	Raw rows → eligible timing rows → filtered pivot rows → site joins → ranking cohort. Preserve a reason for every exclusion and unmatched row.
Duration semantics	Opening, task start, resolution, scan and posting are different timestamps. Verify unit, zone, clock synchronization, boundary rules and reopened-event behavior.
Decision-critical tail	Audit negative, zero, exact-cutoff and missing durations separately. Validate the flag tail, not only aggregate label agreement.
Independent replay	Reproduce the declared filter and duration rule from one frozen source version. A cached pivot need not equal the visible detail population.

[RC/AI] Let N be negative-duration records, E all excluded records and R retained negatives in one nested cohort. Then $R \geq \max(0, N-E)$. If the disputed timing-error category actually denotes negatives and versions align, the registered counts imply a positive bound inconsistent with a zero claim. Alternative category meanings, versions or exclusions may reconcile it; the original label alone does not establish negativity.

[AI] Removing malformed rows can change every site numerator, denominator and standardization reference. Report the rank changes and missingness by process and site; do not treat exclusion as an automatically unbiased repair. A negative interval is a data problem, not exceptionally efficient or improper work.

[PD] The release gate belongs to the data owner and Quality reviewer: source/version agreement, explicit invalid states, linked event identities and independent tail evidence. Do not reverse-engineer a reassuring population by selecting whichever total agrees with the desired conclusion.

Basis: source filter/window assertions and conditional conservation bound [HR/RC]; event-log quality methods [AI]. Historical raw counts and source identifiers remain private.

TECHNICAL ANALYSIS

A within-outcome comparison can reverse the true effect

[AI] Adjustment for process can address one mix difference. Restricting to a booked outcome or later disposition does not make the remaining comparison selection-free. Such variables may be consequences of both workflow choices and severity.

Synthetic F	Severity C	Booked B	Burden Y
0	0	0	0.0
0	1	1	1.0
1	0	1	0.2
1	1	1	1.2

[AI/RC] All four states have probability one quarter. F and C are independent; B=1 if F or C; Y=C+0.2F. Setting F from zero to one increases Y by 0.2 at either severity. Unconditional mean burden rises from 0.5 to 0.7.

[AI/RC] Among booked B=1 cases, F=0 requires C=1, so mean burden is 1.0. For F=1, both severities remain and mean burden is 0.7. The selected contrast is therefore -0.3 despite a true +0.2 effect. Conditioning opens F→B←C. This is a hypothetical possibility proof, not a fitted explanation or allegation about the historical workforce.

[HR/AI] The registered lower-cost fast group weakens the original descriptive cost narrative. Retaining that contrary evidence and withholding a dollar entitlement are sound decisions. Neither the inverse cost pattern nor a larger partial correlation after conditioning identifies the causal effect in either direction.

[AI] Disposal can be a mediator, selection variable or imperfect outcome label. Removing it may remove part of the intervention pathway; conditioning on it may select severity. A duration-band ladder is not a dose-response experiment, and a nonmonotonic ladder cannot distinguish efficient routing, case mix, clock defects or omitted inspection.

[PD] Specify an actual workflow intervention and pretreatment confounders before choosing adjustment variables. Use original eligible cohorts for the total effect. Do not describe administrative category membership as a sufficient control set without a defensible causal structure.

Basis: exact four-state counterexample [AI/RC]; Hernán et al.2004; source post-outcome analyses [HR]. No historical causal coefficient is inferred.

TECHNICAL ANALYSIS

Treat the rank as an audit preference, not an attributable defect rate

[HR/AI] The source combined flag rate, flag count and other site measures, revised a weight after inspecting the list, and retained companion views. These are choices about scarce audit attention. They do not calibrate a probability or identify preventable customer burden.

Quantity	Interpretation and boundary
Flag rate [HR/AI]	Share of declared workflow events meeting the timing rule. Its denominator excludes a later status, and its meaning depends on valid clocks and process mix.
Composite score [HR/AI]	A weighted, transformed ranking index. Its value depends on reference sites, clipping, weights, eligibility and missing joins. It is not a diagnostic posterior.
Outcome rate × flag share [AI]	Different shipment and workflow-event denominators do not form an attributable rate. A joint risk requires a common population and the appropriate conditional probability.
Daily median outcome rate [AI]	Describes a typical day under that definition. It differs from the pooled ratio of outcome counts to exposure; changing window alignment can also change the estimand.

[AI] Near-zero Pearson correlation establishes little linear association in the selected joined snapshot. It does not establish that a KPI is structurally blind, that the new flag is independent of measurement error, or that a nonlinear or subgroup relation is absent.

[PD] Freeze the audit objective and scoring procedure before validation. Compare rate, volume and expected verified value at equal capacity. Keep low-volume sites in an uncertainty-aware review route; a sample floor does not make their true risk zero.

[PD] Recompute ranks under timestamp repairs, missing-data bounds, alternate reference sets and plausible weights. Report near-cutoff bands and held-out decision utility. A weight tuned after observing an inconvenient rank cannot be evaluated as an independent success on the same list.

[AI/PD] Campbell/Goodhart motivate checking whether the target changes recording or pushes work into another workflow. OECD/JRC composite guidance motivates explicit normalization and sensitivity. Neither literature supplies the correct weight or a validated misuse label for this case.

Basis: registered score revision and joining rules [HR]; construct/aggregation analysis [AI]; proposed ranking validation [PD]. No public source weights, site ranks or recoverable rate anchors are shown.

TECHNICAL ANALYSIS

Verify unflagged episodes as well as the cases the screen selects

[PD] Define the reference target as a required procedural step or suitable disposition under an approved standard. Legitimate fast handling, verified deviation, malformed timing and insufficient evidence are separate states. Intent is outside the detector’s claim.

Measure	Prospective definition
Positive predictive value	Verified procedural deviations among valid flagged episodes in the declared population.
Sensitivity / specificity	Flag behavior among independently verified deviations / conforming episodes, using a reference sample from both flag states.
Audit-policy yield	Verified actionable deviations found per fixed audit slot. This is an allocation result, not whole-population diagnostic accuracy.
Missing truth	Selected but unresolved reference records remain explicit. Report identification bounds and failure reasons rather than counting them as conforming.

[PD] Draw a random core from all eligible episodes, stratifying by flag, process, node type and workflow version. Oversample rare timing failures with known inclusion probabilities. For verified record i , use its design weight $1/\pi_i$ in confusion-cell totals; weighting does not correct reference nonresponse automatically.

[PD] A proposed core of 240 independent histories gives a worst-case raw-agreement screen of about ± 6.3 percentage points. This is not a kappa guarantee: prevalence, site clustering and unequal weights require pilot-based precision simulation. Use independent blinded coders, a versioned codebook, adjudication and intervals for agreement, kappa and diagnostic measures.

[AI] Verifying only high-ranked flags creates selective labels. A good PPV in that sample does not establish sensitivity elsewhere. Reusing the scored timestamp as the reference also risks incorporating the test into its own definition of truth.

[PD] Keep calibration separate from policy remediation, or provide identical blinded research verification and a symmetric urgent-safety route in both arms. Reference-quality failures block release even when the numerical classification model looks favorable.

Basis: proposed measurement contract [PD]; STARD2015; Begg–Greenes1983; Lakkaraju et al.2017; Cohen1960. Clinical reporting principles are applied by analogy, not as a clinical or personnel classification claim.

TECHNICAL ANALYSIS

Give every original package a defined observation opportunity

[HR/AI] The source's within-extract recurrence comparison does not provide equal follow-up for all first touches. A later first event has less time to recur before extraction ends, and disposition can change whether another event remains observable.

Identity, censoring and process variation

[PD] Reconcile event multiplicity, unique physical entities, eligible first touches and every exclusion. Freeze the index event and a common follow-up horizon. Distinguish a reopened record, another exception, disposal, return, delivery and customer damage; these are different events.

[AI] No observed repeat can mean resolution, diversion, missing linkage or insufficient time at risk. A survival analysis requires event/censoring times and defensible observation assumptions. Disposal and return can be competing events, not independent censoring. Neither a short-window nonrecurrence count nor a model extrapolation establishes ultimate safety.

[HR/RC/AI] The private audit contains 31 mechanical screens: 12 ordered category profiles, 14 cross-group comparisons and five artifact/basis comparisons. None supplies an actual operating time series or eligible Laney panel. A date inventory does not supply daily observations. No arithmetic point or moving range crosses its provisional limits; this cannot establish stability.

[PD] For compatible ordered observations, tentative XmR uses $MR_t = |x_t - x_{(t-1)}|$, $\text{mean}(x) \pm 2.66 \times \text{mean}(MR)$, and MR upper $3.267 \times \text{mean}(MR)$. Duration bands, process categories and ranked sites are not time series. Do not fabricate daily points or counts from summary rates.

Selection and moderation

[HR/AI] The selected tail depends on an event-volume gate and a tuned score. The record supplies no repeated untreated ranking reliability. If an initial extreme-score excess regresses with hypothetical reliability 0.30/0.60/0.90, expected untreated reversion is 70%/40%/10% of that excess. These are sensitivity assumptions, not estimated fractions of the observed finding.

[PD] Freeze selected membership and track selected and nonselected eligible domains. Analyze process, business line, node technology, task severity and clock-quality moderators with uncertainty. Different workflows can have different legitimate duration distributions; a universal cutoff requires invariance evidence.

[AI/PD] Within-limit arithmetic does not prove stability; a detected shift does not prove treatment causality. A future comparison needs compatible phases, aligned populations and exposure before it can support an operating conclusion.

Basis: source recurrence/exclusion limits [HR/AI]; Kaplan–Meier1958; NIST; Laney; Barnett. The final source inventory and all raw coordinates remain private; no historical correction is estimated.

TECHNICAL ANALYSIS

Credit the diagnostic work while keeping its implementation boundary

[HR/AI] The register attributes filter investigation, detector design, score revision and the challenge response to the case author. Separate operating owners held reporting, adoption and corrective-control responsibilities. The source does not establish deployment of those recommendations.

Contribution claim	Process-tracing grade and boundary
Analytical authorship [HR/AI]	Provisional registry hoop support for the workflow, questions and outputs. Source recovery and provenance determine whether a stronger claim is justified.
Revision and restraint [HR/AI]	Provisional hoop support for retaining contrary outcomes and declining a financial claim. That is a recorded analytical decision, not realized cost avoidance.
Validation harness [HR/AI]	Reported scripts and agent checks support a repeatability intention. They do not constitute independent reference truth or external peer review.
Operating/customer effect [HR/AI]	Unestablished. A recommended audit or second check is not implementation, and implementation alone would not identify a mature effect.

Consolidated limitations

- [AI] Original artifacts were not recovered; contribution rests on register transcriptions. Populations, timing fields, filter states, windows and first-touch exclusions require reconciliation. Same-category or disposal-adjusted comparisons retain selection risks. No independent historical classification study establishes that flagged episodes omitted a required step.
- [AI] No mature counterfactual establishes preventable customer loss, remedy effectiveness or cash release. Selected site rankings do not transport automatically across workflows or future versions. Missing delivery evidence establishes neither completion nor failure outside the supplied corpus.
- [PD] Preserve source hashes, query/filter definitions, event-to-entity crosswalks, analysis versions, selected denominators, unknown states, alternative interpretations and executable checks. Recovering a workbook can authenticate its rows without authenticating an interpretation of human behavior.
- [PD] Publish aggregate anonymous conclusions. Retain raw rows, identifiers, names, dates, ranks and transformation keys privately. The defensible public claim is building and challenging an operational diagnostic, then specifying its validation and release conditions.

Basis: selected role and governance records [HR]; provisional process tracing [AI]; Collier. Original recovery status is maintained in the private manifest and must not be inferred from the completeness of a summary.

TECHNICAL ANALYSIS

Test audit allocation at the same capacity and count the remedial work

[PD] Randomize independent workflow domains to screen-prioritized or uniform audit allocation, blocked by a prespecified baseline risk stratum. Both arms use the same audit capacity, independent adjudication and approved remedy. No person-level decision or operational grant is automated.

[PD] The proposed model uses a frozen pool of 100 eligible episodes and ten audit slots per domain. Declare tie breaking and how unused flag capacity is filled. Keep every original episode in the ITT, including unaudited events, no defect found, failed correction and unresolved follow-up. Shared staff or remedy ownership must remain within the assignment domain.

[PD] Primary benefit is usual-minus-prioritized mean resource burden from independently verified recurring or unresolved protocol defects at original episode + 28 days, audit work and corrective effort. Repair does not restart follow-up. Charge fixed screening cost once outside the variable contrast. Detection yield is secondary: a confirmed defect creates value only through an effective remedy that covers its cost.

[PD] The hypothetical effect path varies common remedy efficacy in both arms. It does not reinterpret a sensitivity parameter as measured adherence. Derive the domain law from finite flags, fixed audit slots and residual domain heterogeneity; evaluate economic-boundary variance as well as the favorable alternative.

[PD] Compare MDE with break-even in identical reference units. Pilot baseline defect prevalence, diagnostic validity, remedy efficacy, dependence, elapsed follow-up and capacity. If independent domains cannot be recruited, redesign and re-power; correlated rows do not replace clusters. Numerical planning does not establish field feasibility.

[PD] Independently size false corrective actions, serious customer harm, queue displacement and missing-reference safeguards. Research adjudication must not silently improve one comparator arm; calibrate separately or apply an identical reference protocol with a symmetric urgent-safety exception.

[PD] A later staggered field rollout targets group-time ATT against not-yet-treated domains under no anticipation, support and interference controls. Event-study and equivalence diagnostics inform parallel-trends assumptions; nonsignificance does not prove them. Mature customer valuation and power require new data, not the diagnostic snapshot.

Basis: proposed domain-randomized audit policy and synthetic model [PD]; selective-label and decision-value methods. No executed trial, validated remedy effect or historical saving is asserted.

TECHNICAL ANALYSIS

Make the workflow contract earn its release through four tests

[AI/PD] WVC turns a suspicious-looking timing distribution into falsifiable measurement and decision claims. It preserves the source’s useful analytical restraint while refusing to equate a ranked signal with truth, responsibility or preventable loss.

Proposition [PD]	Evidence that challenges it [PD]
P1: Valid clock and cohort rules improve defect discrimination.	A blinded held-out comparison with precise absence of the prespecified useful gain, or material unresolved tail errors, challenges release.
P2: Independent reference calibration transports across declared workflows.	Prespecified process/version strata show material miscalibration or reference failure after appropriate weighting and uncertainty.
P3: Prioritization improves net outcomes under equal audit capacity.	An upper net-benefit bound below zero contradicts value; a wide interval is inconclusive. A failed harm guardrail blocks scale independently.
P4: Verified remedy delivery changes repeat defects beyond ticket closure.	An adequately powered controlled follow-up shows no useful durable gain, or the proposed readback fails to confirm the remedy reached the workflow.

[AI] Selection-bias theory makes the conditioning set part of the causal question. Selective-label research distinguishes what is unobserved because of earlier choices from what is absent. Event-log quality and diagnostic reporting add separate source, index-test and reference-standard obligations. A clean transform cannot substitute for a valid target.

[AI] Decision-value research explains why accuracy or yield alone cannot choose an audit policy. Costs of unnecessary review, ineffective correction and displaced work matter. The synthetic economics is a proposed decision model; independent information value purchases a better choice and must not be added to the same avoided operating loss.

[AI] Deming discourages reactions to unqualified variation. Campbell/Goodhart motivate checking substitution into other workflows after a timing target is introduced. Composite-indicator guidance requires transparent weights and sensitivity, while Ittner and colleagues connect measurement to decisions rather than metric volume.

[AI/PD] Tucker and Edmondson distinguish repeated local workarounds from learning that removes recurrent causes. A closed audit ticket is therefore insufficient: the owner must verify implementation and later recurrence. None of these propositions is historically established by the source snapshot.

Basis: proposed WVC propositions [PD]; Hernán; Lakkaraju; Suriadi; STARD; Vickers–Elkin; Deming; Campbell/Goodhart; OECD/JRC; Ittner; Tucker–Edmondson. No individual misconduct conclusion is made.

PROSPECTIVE MODEL / SIGNAL VALIDATION AND AUDIT ALLOCATION

Use a screen to allocate review, never to establish misconduct

[PD] Every number on these four pages is hypothetical. Validate definitions on independently sampled flagged and unflagged episodes before a trial. Known verification probabilities, blinded adjudication and unresolved-label bounds are required; speed is not intent.

Fixed design	Audit-policy comparison
Original population	100 eligible workflow episodes per independent domain; 10 review slots. Randomize domains within preassignment risk strata to uniform or signal-prioritized audit.
Truth and allocation	True defect means required-step omission or unsuitable disposition. Prioritized audits sample flagged episodes; if fewer than 10 exist, fill from unflagged episodes. No personnel consequence follows a flag.
Follow-up	The same approved repair is attempted for every reference-confirmed defect. Independently assess recurring/unresolved protocol defect at original episode + 28 days; retain unaudited episodes and failed repairs.

[PD] Domains contain 4 or 12 true defects with equal overall frequency. Sensitivity 0.75 and false-positive probability 0.15 imply population PPV 30.30%. Most flags are not confirmed defects; this does not itself show that priority auditing lacks value.

REFERENCE COLLECTION IS NOT THE INTERVENTION

[PD] Validation and outcome reference work remain separate from allocation audits, with symmetric urgent-safety exceptions. The numeric law assumes ideal reference truth; pilot misclassification, unresolved labels and transport before adopting its yield estimates.

[PD] Residual stratum misclassification is 10%, stressed to 25%; within-domain clustering is retained. Follow-up is a proposed protocol horizon, not validated customer-outcome maturity.

PROSPECTIVE MODEL / EXACT MECHANISM, CONDITIONAL POWER

Power the value of an audit policy on the whole original cohort

[PD] Exact binomial flag counts and hypergeometric fixed-capacity selection give audit-allocation moments. They feed an explicitly approximate Gaussian planning law; the model does not claim exact field power.

Design	Per arm	MDE (index)	Break-even detection	Full-effect clearance
Initial	20	> full path	5.0%	9.9%
Base candidate	2860	2.96	80.3%	>99.9%
Same size / stress	2860	3.70	60.6%	>99.9%
Revised / stress	4920	2.82	83.8%	>99.9%

[PD] A lower 97.5% bound on variable-benefit ITT must exceed screening/support cost 2.96 index points. Also require 80% two-sided 5% detection power at that break-even. Minimum recruitment is 2844 domains per arm under base strata, 4462 under weaker strata.

[PD] Repair reduces a true defect’s repeat probability from 0.60 to $0.60-0.40\lambda$ in both policies. λ is common remedy-efficacy sensitivity, not compliance. Variance is maximized over λ in $[0,1]$; finite-process simulations regenerate outcomes at the zero and economic boundaries.

A NUMERICAL PASS IS NOT A FEASIBLE FIELD PLAN

[PD] The revised candidate requires 9,840 independent domains and 984,000 original episodes. Recruitment is unestablished; do not recommend this rollout. Fund bounded detector/reference and variance/remedy calibration first. The causal investment decision remains unresolved; no customer or personnel assurance follows.

[PD/RC] MDE and cost are index points; usual protocol-defect resource burden=100. Allocation effects use all original episodes, not only audited, repaired or concession-bearing records.

PROSPECTIVE MODEL / CONFIRMED DEFECTS AND FULL REPAIR COST

Value verified repeat-defect reduction and debit every attempted repair

[PD] A repeat defect costs 20 hypothetical resource units; each confirmed repair attempt costs 2 whether it succeeds or fails. Both policies spend 10 on their fixed audit capacity. Equal review cost remains in each arm’s burden even though it cancels from the contrast.

[PD] Under the full assumed remedy effect, priority review adds 2.147 confirmed repairs and prevents 0.859 repeat defects per 100 original episodes in expectation. These are model contrasts, not observed corrections. Usual burden is 101.2 resource units, indexed to 100.

Avoidable / non-overlap	Cost index=1.98	Cost index=2.96	Cost index=3.95
0.50 / 0.80	+0.57	-0.42	-1.41
0.75 / 0.90	+5.24	+4.25	+3.26
1.00 / 1.00	+10.75	+9.76	+8.78

[PD] $Net_raw = a \cdot h \cdot 20 \cdot \text{repeat-defect ITT reduction} - 2 \cdot \text{extra repairs} - K_raw$. $Net_index = 100 \cdot Net_raw / 101.2$. $K_raw = 3$ per original 100-episode pool allocates screening 1.2, coordination 0.8 and recurring verification 1.0; research-only adjudication is separate.

YIELD IS NOT VALUE WITHOUT A WORKING REMEDY

[PD] More confirmed defects can create more repair cost without enough repeat reduction. Discount avoided repeat-defect value only, never new repair or screening expense. Field customer valuation requires a separate mature outcome effect and costs; this protocol endpoint cannot supply them.

[PD/RC] Net grid uses index points, not money. Repair efficacy and follow-up are hypothetical. Retain delays, unsafe dispositions and unresolved reference states in separate acceptance gates.

PROSPECTIVE MODEL / INFORMATION BEFORE INTERVENTION

Distinguish a broken definition from a weak remedy before scaling

[PD] This separate utility model values a better choice among a bounded audit-allocation pilot, measurement repair and remedy testing or continued hold. Its information value does not require the proposed detector or remedy to be effective.

Action / utility	Usable + effective	Invalid basis	Other / weak remedy
Controlled audit pilot	15	-10	-9
Repair measurement	-4	13	-4
Test remedy / hold	1	-5	12

[PD] Priors are 0.40/0.35/0.25. Invalid definitions/reference take priority; among valid cases, usable and effective requires evidence of both. The remainder includes weak remedy or unresolved applicability. Signals identify the state with probability q; each alternative has probability (1-q)/2.

Signal quality q	EVSI	Evidence cost	Net EVSI
0.50	2.4000	0.50	1.9000
0.75	7.0000	1.50	5.5000
0.90	9.7600	3.00	6.7600

AN AUDIT PRIORITY IS NOT AN ALLEGATION

[PD] Enumeration of 27 signal-action policies gives prior value 1.95 and EVPI 11.60. Utilities include later action expense; acquisition cost is subtracted once. Do not add this utility to operating savings or use it to justify consequences for an individual.

[PD/RC] All state probabilities, signal accuracies, utilities and costs are assumed. This is decision-information value, not a historical saving or evidence of worker intent.

Reproduce the population bridge before interpreting the audit list

[AI] This is an evidence-reconciliation and proposed evaluation case. The historical contribution is an analytical detector and its challenge response. The new work tests the population logic, selection assumptions and evidence needed before a signal can direct corrective action.

Claim class	Meaning and replication boundary
[HR] Historical record	A statement or reported action in the selected register. An earlier source’s assertion that it verified a workbook is not the same as recovering and auditing that workbook here.
[RC] Recalculated	Count conservation, cohort and version reconciliation, tentative process limits and explicitly hypothetical model results. Correct totals do not establish that an event label measures a physical task.
[AI] Analytical inference	A conclusion tied to named assumptions: incompatible nested counts, post-action selection, unstable ranking, unequal follow-up or a missing bridge from flags to preventable defects.
[PD] Proposed design	The Workflow Verification Contract, independent reference audit, equal-capacity policy evaluation, economic model and information-value analysis. None is a historical deployment.

Reproduction sequence

[PD] Verify the source version and selected cells. Reconcile raw records, valid windows, dispositions, site joins and unique identifiers as separate states. Preserve every competing timestamp and filter definition; do not treat source confidence as an independent validation.

[PD] Recompute the retained-error bound and selected-subgroup counterexample. Independently rebuild the finite audit-allocation law, reference assumptions, repeat-defect outcomes, repair costs and sample-size decision. Reprice any changed efficacy, capacity, truth or cost assumption before interpreting value.

THE DECISION THE CASE CHANGES

[AI/PD] A ranked timing anomaly is an invitation to verify a process. Funding depends on verified, preventable defects and the value of the chosen audit policy. Intent, attributable customer loss and realized savings require evidence beyond the screen.

Public source conclusions are anonymized; numeric illustrations are explicitly hypothetical. Confidential raw values, identity mappings, source paths and reproduction code are private.

RESEARCH REFERENCES / VALIDATION AND SELECTION

Validate the reference standard and preserve the selection mechanism

Primary research and official methods inform signal validation, selection and evaluation. They do not authenticate unavailable project sources or establish the detector's accuracy.

Hernán, M. A., Hernández-Díaz, S., & Robins, J. M. (2004). A Structural Approach to Selection Bias. *Epidemiology*, 15, 615–625. DOI 10.1097/01.ede.0000135174.63482.43. Full original reproduction. Conditioning on a common effect can create an association. The case distinguishes process adjustment from selection on bookings or disposition.

Lakkaraju, H., et al. (2017). The Selective Labels Problem: Evaluating Algorithmic Predictions in the Presence of Unobservables. *KDD*, 275–284. DOI 10.1145/3097983.3098066. Full author paper. Labels observed only after a selection decision can bias evaluation. The proposed reference audit includes flagged and unflagged events rather than validating only investigated records.

Suriadi, S., Andrews, R., ter Hofstede, A. H. M., & Wynn, M. T. (2017). Event Log Imperfection Patterns for Process Mining: Towards a Systematic Approach to Cleaning Event Logs. *Information Systems*, 64, 132–150. DOI 10.1016/j.is.2016.07.011. Institutional primary abstract. Event identity, completeness and timestamp semantics require explicit quality checks before process interpretation.

Bossuyt, P. M., et al. (2015). STARD 2015: An Updated List of Essential Items for Reporting Diagnostic Accuracy Studies. *BMJ*, 351:h5527. DOI 10.1136/bmj.h5527. Full original guideline. The case adapts transparent eligibility, index-test, reference-standard and flow reporting. It does not claim clinical validation or compliance certification.

Vickers, A. J., & Elkin, E. B. (2006). Decision Curve Analysis: A Novel Method for Evaluating Prediction Models. *Medical Decision Making*, 26, 565–574. DOI 10.1177/0272989X06295361. Indexed primary manuscript; direct opening restricted. Useful prediction depends on the action and consequences of errors. The case uses its own fixed-capacity decision model, not a borrowed clinical threshold.

Kaplan, E. L., & Meier, P. (1958). Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Association*, 53, 457–481. DOI 10.1080/01621459.1958.10501452. Full original reproduction. Incomplete follow-up must be explicit. The source lacks the event-time data and assumptions required for a fitted survival or competing-risk analysis.

Primary-source access scope and metadata are recorded privately. Methodological analogies do not import clinical thresholds, external effect sizes or validation claims.

RESEARCH REFERENCES / UNCERTAINTY AND EVALUATION

Test process signals and audit policies on their actual decision units

Primary research and official methods inform signal validation, selection and evaluation. They do not authenticate unavailable project sources or establish the detector's accuracy.

NIST/SEMATECH. e-Handbook of Statistical Methods, section 6.3.2.2: Individuals Control Charts. Official method. Moving-range limits require ordered, comparable observations. Ranked sites, speed categories and revised score coefficients are not a calendar process, even when a mechanical screen can be calculated.

Laney, D. B. (2002). Improved Control Charts for Attributes. *Quality Engineering*, 14(4), 531–537. Publisher abstract. Large counts do not remove clustering or incompatible denominators. The case retains eligibility checks before treating an attribute chart as process evidence.

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20, 37–46. Primary publisher record. Independent coding reports confusion matrices, raw agreement and kappa with uncertainty. Agreement with an existing duration label is different from agreement on a verified protocol defect.

Barnett, A. G., van der Pols, J. C., & Dobson, A. J. (2005). Regression to the Mean: What It Is and How to Deal with It. *International Journal of Epidemiology*, 34, 215–220. Primary abstract. A selected extreme audit cohort can move toward average without corrective action. Repeat untreated rankings or explicit sensitivity assumptions are needed.

Hudgens, M. G., & Halloran, M. E. (2008). Toward Causal Inference With Interference. *Journal of the American Statistical Association*, 103(482), 832–842. Primary manuscript. Shared staff, queues and corrective practices can affect comparison units. Randomization must match the operational dependency boundary or account for spillovers.

Callaway, B., & Sant'Anna, P. H. C. (2021). Difference-in-Differences with Multiple Time Periods. *Journal of Econometrics*, 225, 200–230. Author manuscript. Group-time effects inform a staggered alternative with explicit treatment timing and comparison populations. Its assumptions and power do not follow automatically from the randomized model.

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. The case values information through a separate finite-state decision model and keeps it distinct from the operating outcome benefit.

Primary-source access scope and metadata are recorded privately. Methodological analogies do not import clinical thresholds, external effect sizes or validation claims.

RESEARCH REFERENCES / INDICATORS AND OPERATING LEARNING

Connect a diagnostic score to a governed learning process

Primary research and official methods inform signal validation, selection and evaluation. They do not authenticate unavailable project sources or establish the detector's accuracy.

Collier, D. (2011). Understanding Process Tracing. *PS: Political Science & Politics*, 44, 823–830. Primary publication. Analytical authorship, score revision, reporting ownership and corrective implementation need separate contribution evidence. Multiple derivative summaries do not create independent corroboration.

Tucker, A. L., & Edmondson, A. C. (2003). Why Hospitals Don't Learn from Failures. *California Management Review*, 45(2), 55–72. Author manuscript. Resolving an immediate exception differs from correcting recurrent causes. An audit list creates learning only when verified findings produce accountable action and later feedback.

The W. Edwards Deming Institute. The Funnel Experiment. Official exposition. Reactive adjustment to noise can worsen a process. This case requires valid populations and stable definitions before interpreting movement; a suspicion score does not itself provide a process-control limit.

Campbell, D. T. (1976). Assessing the Impact of Planned Social Change. Original institutional record. Consequential indicators can alter behavior and reporting. WVC separates the process score from independent outcome evidence and preserves unflagged records in validation. No historical intent or manipulation is inferred.

Goodhart, C. A. E. Problems of Monetary Management: The U.K. Experience. 1975 conference; revised volume listed as 1976 in the RBA archive. A target can change the relationship between a measure and its underlying construct. Closure speed, recorded disposition and physical resolution therefore require distinct definitions.

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. A score's operating value depends on the decisions it supports. More elaborate weighting does not independently validate a diagnostic or identify a causal benefit.

OECD / European Commission JRC (2008). Handbook on Constructing Composite Indicators. Official handbook. Normalization, weighting, aggregation and sensitivity shape rankings. This case treats the score as an audit-allocation policy and preserves near-cutoff uncertainty rather than presenting rank as a calibrated probability.

Primary-source access scope and metadata are recorded privately. Methodological analogies do not import clinical thresholds, external effect sizes or validation claims.

FINAL REFEREE REPORT / INTERNAL AGENT-ASSISTED REVIEW

Accept the diagnostic case within its unresolved validation boundaries

[AI] This is author-led, agent-assisted review. It is not external peer review, client approval or a consulting-firm ranking. The supported contribution is a documented diagnostic, a reconciliation of its evidence and a proposed audit-policy evaluation.

Major issue	Disposition in the revised case
Population and timing	Made the retained-negative bound conditional on the error-category meaning and aligned populations. Preserved competing timestamp and exclusion versions; aggregate label agreement does not prove tail validity.
Selection and entity	Separated events, unique identifiers and the first-touch analysis population. Post-action bucket and disposition conditioning remain potential selection mechanisms; short follow-up does not establish safety.
Ranking and accuracy	Kept the composite as a capacity-allocation rule. Weights, reference distribution and cutoff require frozen out-of-sample validation; flag status does not establish a protocol defect or intent.
Evaluation and economics	Compared audit policies at equal capacity and retained original eligible episodes. Value depends on independently verified outcomes, repair effectiveness and full incremental cost; no historical cost leak or saving is asserted.
Contribution and review	Bounded authorship, challenge response and financial restraint separately from reporting and corrective ownership. Source, numerical and actual-page reviews are preserved privately.

Partner critique: can a COO act on page one?

[AI] Yes: commission the event-universe reconciliation and independent reference audit before assigning corrective action from the screen. Data owns the count bridge; Quality owns defect adjudication; Operations owns capacity and authorized remedies.

Minor revisions and remaining limits

[AI] Replace misconduct labels with screening language; preserve disputed identity and role claims privately. Calibration must establish reference quality, recruitment, follow-up and costs. A passed simulation is not field validation.

REVIEW DISPOSITION / PASS WITHIN STATED SCOPE

[AI] The case supports an analytical diagnostic and a reproducible proposed evaluation. It establishes no individual misconduct, validated detector accuracy, implemented corrective control or realized benefit.

Scope: internal evidence, scientific, numerical and executive review. The review verdict is limited to the claims supported by the supplied record and explicit hypothetical analysis.