

KPI Definition & Reporting Reliability

Connect every reported value to a verified definition, population and source version

An anonymized operating case on an executive reporting application whose KPI names, filters and aggregations produced conflicting answers. The register describes identifier conventions, arithmetic checks and explicit unavailable states for unsupported values and attribution.

PROJECT CONTEXT / KPI GOVERNANCE & DECISION ASSURANCE

BLUF / BOTTOM LINE UP FRONT

Pin each KPI to its definition and population.
Make verification a condition of release.

ANALYTICAL FINDING

A recorded population-risk conclusion fails its own arithmetic: assuming matched nested populations reverses the claimed ordering. Because the source also mixes scope labels, subgroup comparisons must remain unresolved until their lineage is reconciled.

CONTRIBUTION IN THE SOURCE RECORD

Directed KPI mapping, diagnostic hypotheses and release rules. Provisional registry hoop support; source ownership and app implementation remained external.

EVIDENCE STATUS

Measurement and design case; originals unrecovered. Live validation, broad adoption, customer impact and savings are unestablished.

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

EXECUTIVE ABSTRACT / DECISION AND OWNERSHIP

Require a verifiable KPI binding before an operating decision

[AI] A KPI name can remain familiar while its population, denominator, source version or meaning changes. The proposed Metric Identity and Release Contract (MIRC) makes that full binding an explicit condition of publication.

Abstract element	Decision-relevant synthesis
Context [HR]	An executive reporting application returned conflicting values, used ambiguous identifiers and displayed unsupported breakdowns. Later access failure prevented fresh validation through the recorded connector.
Objective [PD]	Make each reported KPI reproducible for its declared population and date, and prevent unavailable evidence from being presented as a current result or a causal explanation.
Approach [HR/PD]	The project recorded identifier mappings, scoped checks, arithmetic rules and unavailable states. This case audits their assumptions and proposes a versioned release contract.
Findings [RC/AI]	The reported late-defect correction removes about an 85% overstatement, not an 85% operating improvement. A subgroup-risk explanation contradicts its assumed arithmetic; unmatched scope prevents a definitive risk ordering.
Decision value [PD]	Test the cost of valid reuse, authorized rebinding, necessary holds and incorrect releases across repeated refreshes. Price information separately from any customer or operating benefit.
Implications [PD]	The reporting owner accepts the KPI meaning; metric owners resolve definition changes; the data steward versions the binding; an independent reviewer verifies release evidence.

EXECUTIVE ACTION AND THRESHOLD

[PD] Start with a bounded calibration pilot. Release only when identity, scope, denominator, transform and freshness checks pass. Before scaling, establish recruitment feasibility and require the lower confidence bound on resource benefit to exceed matched-horizon cost, with separate false-release and availability guardrails.

Basis: S1 project register and source versions. [HR] historical record; [RC] recalculated; [AI] analytical inference; [PD] proposed design. No historical causal benefit is identified.

EXHIBIT 1 / SEMANTIC IDENTITY

An exact identifier does not establish a valid current KPI

[HR] The register describes replacing ambiguous lookup with explicit identifiers and correcting a spelling linked to the wrong or nonexistent series. [AI] That correction is necessary but does not establish every other part of the metric contract.

IDENTITY IS A VERSIONED SEMANTIC BINDING
Proposed MIRC / an identifier alone supplies only one part of the contract

MEANING Definition + version	ELIGIBILITY Population + exposure	TIME + UNITS Window + transform	EVIDENCE Verification + owner
--	---	---	---

A familiar name can point to the wrong construct or a stale definition.
Approved aliases are permitted; unsupported equivalence is not.

Source: S1 D02–D03/D08. The contract is proposed; mappings do not certify current values.

Binding element	Release evidence [PD]
Meaning and version	Business construct, authoritative identifier, formula version and effective date; approved aliases retain their semantic relationship.
Population and exposure	Eligible geography, business line, node type, numerator grain and metric-specific denominator. A familiar label cannot supply them.
Time and transform	Observation window, completeness, units, multiplier and aggregation rule. Empty responses and raw fractions are not display-ready values.
Verification state	Who verified the binding, against which source, when, and what changes trigger a new check. A historical mapping remains historical.

[AI] Reject distinct constructs that unexpectedly resolve to the same identifier. Do not reject every many-label-to-one-ID mapping: approved aliases can be legitimate. Conversely, different IDs can still measure the same construct or overlapping outcomes.

OPERATING IMPLICATION
[PD] Bind the semantic tuple, preserve known source spellings verbatim, and require owner review when the tuple changes. Record unresolved identity separately from unavailable current values.

Source: S1 D02–D03, D08 and identity/reference records. MIRC is proposed; the existence of a written mapping does not prove current validation or enterprise adoption.

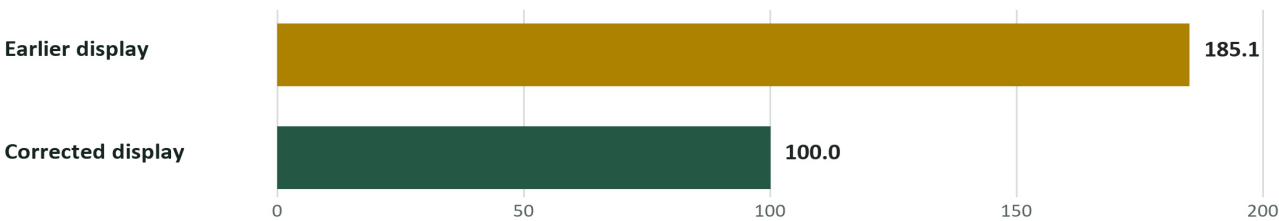
EXHIBIT 2 / REPORTED-VALUE CORRECTION

A large reporting correction is not a reduction in customer defects

[HR] A late-defect value changed when the reporting scope was corrected. [RC] Relative to the corrected display set to 100, the earlier display is approximately 185.1. The observations describe one reporting question, not an operating intervention.

THE CORRECTION CHANGES THE REPORTED VALUE

Historical approximate display comparison / corrected value = 100



85.1% overstatement of the reference = 46.0% decline from the earlier display

Source: S1 M02. A reporting correction does not establish an operating effect.

[RC] A value 85.1% above the corrected reference must fall about 46.0% from its own starting level to reach that reference. Those two percentages use different denominators. Neither estimates fewer late deliveries, a treatment effect or financial savings.

Reconciliation step	What the record supports
Numerator and denominator	Both change across the scope versions. A denominator-only story cannot explain the direction or magnitude of the correction.
Precision	The ledger preserves source displays and recalculation from supplied counts. One component differs at the displayed decimal; hidden fractional units remain a possible explanation.
Acceptance tolerance	A historical match band is a point-in-time comparison rule. It does not establish long-run accuracy, and a result inside the band is not an exact match.

OPERATING IMPLICATION

[PD] Publish the corrected scope and the affected-report notice. Retain the error as a measurement correction; evaluate any later operating action on a separate, mature customer cohort.

Source: S1 M01–M06, D01 and E05–E06. Public index uses corrected reported late-defect value = 100. Approximate source displays govern this comparison; raw inputs remain private.

EXHIBIT 3 / SCOPE AND INTERPRETATION

The recorded risk ordering does not follow from the scope bridge

[HR] The narrative interprets different numerator and denominator reductions as evidence that the retained population has higher risk. [AI] The direction is reversed under the narrative’s own nested-population assumption, while the recorded scope labels prevent that assumption from being accepted as fact.

TEST THE EXPLANATION AS WELL AS THE NUMBER

Conditional inference / the broad scope labels are not reconciled



The source’s higher-retained-risk story does not follow.
Without matched scope, the actual subgroup ordering remains unidentified.

Source: S1 M07–M08. This is an algebraic sensitivity, not an observed subgroup result.

[AI] If A is an affected-unit numerator and N is exposure, retaining a smaller share of A than of N implies a lower retained rate: $(A_r/N_r)/(A_\beta/N_\beta) = (A_r/A_\beta)/(N_r/N_\beta)$. With nonnegative, matched and nested populations, the excluded complement must then have a higher rate than the retained group.

Evidence branch	Decision consequence
Assume matching and nesting	The implied direction contradicts the narrative. The source’s risk-direction statement cannot be carried into an operating recommendation.
Require demonstrated lineage	The broad numerator and denominator carry different geography labels. Subtraction may not produce any coherent excluded population.
Across both branches	Reject proportional scaling as an unvalidated shortcut and hold subgroup targeting until common population, date and exposure definitions are verified.

NEW ANALYTICAL FINDING

[AI] The same reconciliation can expose an error in the explanation without identifying the true subgroup risk. The correction is to withdraw the unsupported ordering, not reverse the operating target on assumed data.

Source: S1 M07–M08 and narrative attribution. Diagram is conditional algebra, not an observed subgroup comparison. Raw and alternative-scope calculations are retained privately.

EXHIBIT 4 / ARITHMETIC CHECKS

A failed invariant can stop release; a passed invariant cannot certify truth

[HR] The project used arithmetic checks to challenge a headline below its component sum and a week-to-date cell below the reported daily range. [AI] Each check is valid only for a defined measurement family and set of assumptions.

ARITHMETIC CHECKS DETECT DEFINED CONTRADICTIONS

Proposed guard family / assumptions must travel with the test

COMPONENTS Disjoint + same exposure	TIME AGGREGATION Same days + positive weights	UNIT RANGE Bounded numerator
---	---	--

A violation requires diagnosis of the result or its assumptions.
A pass does not validate the source, population or current definition.

Source: S1 M14–M16. Common errors can preserve every displayed arithmetic relationship.

Proposed guard	Conditions and boundary
Total ≥ component sum	Nonnegative, disjoint component counts; common exposure, scope, time and version. Overlapping siblings can legitimately exceed a parent and need another check.
Period rate within daily range	The period equals a ratio of sums over the same complete days with positive exposures. Different windows, signed adjustments or missing days invalidate the comparison.
Affected-unit rate within unit range	A unique-unit or bounded fractional numerator is contained within its eligible opportunities. Defect-event counts can exceed units; the label “per million” alone imposes no universal ceiling.

[AI] Two systematically mis-scoped series can agree and satisfy every arithmetic check. A common scale error can preserve both component inequalities and weighted-average bounds. Passing therefore rules out selected contradictions; it does not prove source validity.

OPERATING IMPLICATION

[PD] Record the assumptions behind each guard. Quarantine violations for diagnosis, preserve rejected values and their lineage, and require independent source evidence for a release.

Source: S1 M14–M16 and D05/D07. Mathematical conditions are analytical inference; guard design is proposed. Repeated-event metrics and overlapping driver lists require separate contracts.

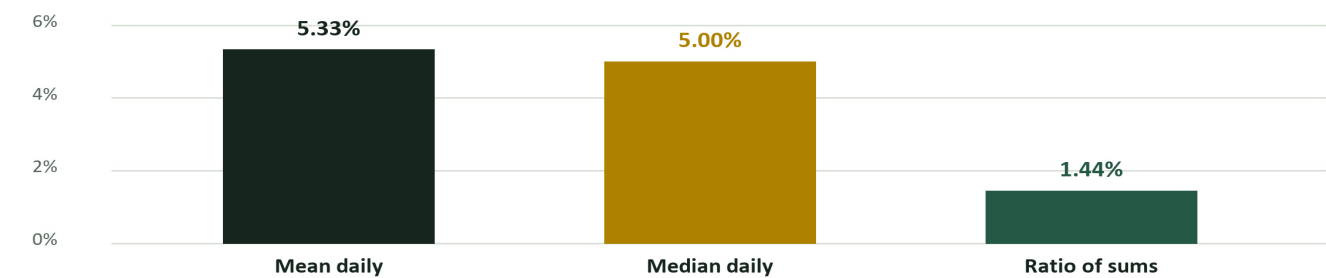
EXHIBIT 5 / AGGREGATION AND ROBUSTNESS

A median daily rate answers a different question from a weekly exposure rate

[HR] The recorded response to a corrupt export was to exclude suspect days and replace the site-level mean with a median. [AI] A median can resist extreme observations while still changing the quantity leaders believe they are comparing.

CHOOSE THE SUMMARY THAT ANSWERS THE DECISION

Synthetic example / period rate differs from the typical-day summary



All three are legitimate summaries of different quantities.

Source: Luna Sol synthetic three-day example. No historical site data are shown.

Synthetic day	Affected units	Eligible units	Daily rate
A	1	10	10.00%
B	5	100	5.00%
C	10	1,000	1.00%

[RC/AI] In this synthetic example the median daily rate is 5.00%, the arithmetic mean is 5.33%, and the period rate is 16/1,110 = 1.44%. The median describes a typical day; the ratio of sums describes exposure-weighted affected units. All three calculations are correct for different questions.

[PD] Use the ratio of valid numerator and exposure sums for the declared period-risk estimand. Retain a median as a separately labeled diagnostic. Validate impossible-value rules against the metric definition; do not delete genuine spikes merely because they are inconvenient.

OPERATING IMPLICATION

[PD] Preserve rejected-day coverage and reconstruct missing or corrupt counts where possible. If exclusion changes the eligible population, publish a qualified window or bounds instead of presenting the remainder as the complete week.

Source: S1 M16 and D07 for the recorded response. All numerical day values above are invented analytical examples; they are not site observations. NIST weighted-mean/location methods inform interpretation.

EXHIBIT 6 / ATTRIBUTION AND ROUTING

A driver list needs semantic validation before it becomes an explanation

[HR] The project distinguished a headline metric from reason quantities and sibling process metrics, and specified unavailable states for unsupported breakdowns. [AI] Neither a plausible pie chart nor a clean keyword audit establishes causal attribution.

A RELATED MEASURE IS NOT AUTOMATICALLY A CAUSAL DRIVER

Analytical interpretation / distinguish each reporting object

OFFICIAL TOTAL Defined parent measure	DRIVER LIST May overlap or differ	ROUTING TERMS Lexical matching evidence
---	---	---

Validate parent coverage, denominators and semantic meaning.
An attractive decomposition cannot fill an unavailable attribution path.

Source: S1 M11–M13/M17, D04/D06/D09. Causal attribution needs separate evidence.

Reporting object	Required interpretation
Sibling process measures	A ranked list may overlap or use different exposures. Preserve the official total and validate disjointness before calling the list a decomposition.
Reason quantities	Shares require a defined common parent and coverage. A quantity-only detail row does not become a defect rate without its own valid exposure.
Metric routing keywords	Shared family stems can be intentional. Duplicate counts and containment checks do not estimate semantic routing precision, recall or cross-row ambiguity.
Unsupported explanations	Use an explicit unresolved state and accountable source owner. Fabricated regional multipliers or missing driver data cannot be repaired by more persuasive labeling.

[HR/AI] The mapping extension reports both row growth and fewer lexical collisions, but source row-count wording and collision dispositions retain unresolved differences. Those audit outputs do not certify correct metric selection in new client or operating questions.

OPERATING IMPLICATION

[PD] Evaluate routing on independently labeled questions, including abstentions and legitimate aliases. Require causal evidence before assigning a corrective action to a listed driver.

Source: S1 M11–M13, M17 and D04/D06/D09. Recorded changes and specified controls are separated from verified adoption. Keyword-audit counts remain private.

EXHIBIT 7 / FRESHNESS AND OWNERSHIP

A cached identity can survive an outage while current validity expires

[HR] A recorded access failure prevented live revalidation, so historical profile files became a spelling reference. [AI] Their usefulness for identity does not establish current values, current formulas or current permission to query.

KEEP IDENTITY MEMORY SEPARATE FROM CURRENT VALIDATION

Proposed lifecycle / historical access failure motivates a bounded fallback



Historical spelling evidence can persist when current values cannot be verified.
Every rebind records the definition change and affected reports.

Source: S1 D08/E03/E11. No current production endpoint was tested for this case.

State or event	Required action [PD]
Authenticated binding	Store the source/version, checked semantic tuple, validation date and allowed use; attach a refresh or change-trigger rule.
Potential definition change	Hold affected values pending owner verification. Separate a renamed alias from a changed construct, denominator or benchmark.
Authorized rebind	Validate against independent expected results, preserve the old version and list affected prior outputs.
Source unavailable	Expose unavailable or historical status. A local utility succeeding does not establish that the remote metric endpoint can return current, usable evidence.

[PD] The dynamic benchmark follows independently governed metric domains across repeated refreshes. Count valid reuse, rebind effort, unjustified holds and wrong-semantic releases at every deadline. Start with calibration: the later power analysis requires a larger independent pool than this record establishes.

OPERATING IMPLICATION

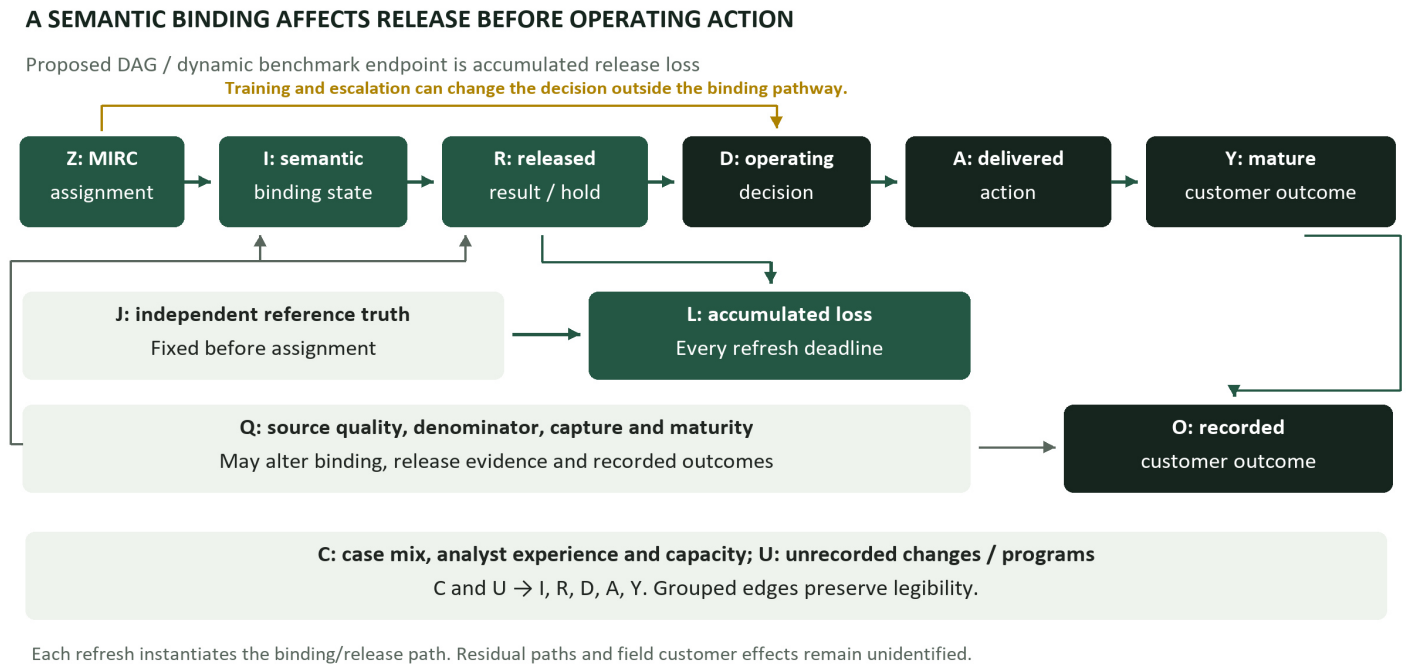
[PD] Assign a metric owner and reporting steward before scaling. A pilot can establish decision reliability under its test conditions; customer outcomes and realized savings require their own evidence.

Source: S1 D08, E03/E11 and governance narrative. Access status is historical to the supplied record; no current production endpoint was tested. MIRC lifecycle and evaluation are proposed.

TECHNICAL ANALYSIS

A valid binding connects the metric to the decision it can support

[AI/PD] The proposed Metric Identity and Release Contract (MIRC) governs the meaning and current validity of a metric binding, including when to reuse it, revalidate it or hold its output.



[PD] Z assigns MIRC availability and orientation. I is the metric binding: concept, version, population, numerator, exposure, clock and transform. I affects released result or hold R; R can change decision D, delivered action A and mature customer outcome Y. Independent truth J and the sequence of release decisions determine benchmark loss L.

[PD] Q represents source capture, denominator and maturity quality; O is the recorded customer outcome. C names case mix, analyst experience and shared capacity; U includes undocumented definitions and concurrent programmes. These can affect binding choices and downstream outcomes. The benchmark would identify Z’s effect on L; neither I’s effect on Y nor A’s effect on Y is thereby identified.

[AI] Historical corrections concern I and R. A different reported value can reflect a corrected definition while operations remain unchanged. Training or escalation may also change D directly, so Z is not automatically an instrument for metric use. Unexplained residuals remain compatible with unresolved capture, definition and operating paths.

Basis: registered design and correction descriptions [HR]; proposed MIRC causal graph [PD]. Customer-effect paths are hypotheses, not established results.

TECHNICAL ANALYSIS

Exact names need a versioned semantic signature

[HR/AI] The register describes mistaken metric selection, incompatible topline and a restricted offline fallback. Exact spelling reduces one failure mode; it does not certify the current meaning of a number.

Contract field [PD]	Release requirement [PD]
Concept and identity	Bind the business question, definition and immutable or versioned reference; distinguish preferred labels, aliases and genuinely different constructs.
Population and quantity	Retain entity, numerator, exposure, eligibility, exclusions, units and supported dimensions. A shared name is not a shared population.
Time and transformation	Freeze observation window, source vintage, maturity convention, scale, aggregation and rounding stage.
Authority and change	Name the definition owner, effective version, revalidation evidence, accepted exceptions and expiry or change trigger.

[AI] A guard that rejects every pair of labels sharing an identifier will reject legitimate aliases. Conversely, separate identifiers may describe equivalent quantities. The relevant test is an explicit equivalence relation over the signature. W3C SKOS distinguishes concepts, labels and mappings; adopting that distinction does not make an unverified local mapping correct.

[HR/AI] The offline profile was bounded to historical spelling support. It cannot establish that a metric remains live, authorized, unchanged or numerically current. A successful lookup or familiar value is also insufficient: the same identifier can retain its name while the denominator or population changes.

[PD] Preserve the old signature and source hash when a change is detected. Rebind only after owner-supported comparison; otherwise publish the precise unresolved state and route it. Equivalent aliases can reuse an authenticated binding. A spelling-only cache can assist discovery but cannot silently supply current values. Revalidation failure must not reset the apparent age of the underlying evidence.

[AI] Hard-coding is therefore a controlled binding choice, not a permanent truth guarantee. Catalog version metadata and cache validation distinguish reference availability from current acceptance; neither establishes business accuracy by itself.

Basis: registered identity and offline-fallback decisions [HR]; W3C SKOS, DCAT and RFC 9111 as scoped design references [AI/PD].

TECHNICAL ANALYSIS

Population and time aggregation define the rate being compared

[AI] A numerical correction becomes interpretable only after its numerator, exposure and time window describe the same quantity. A plausible scale factor cannot supply that missing contract.

[AI] For compatible daily counts N_d and strictly positive exposures E_d , the period rate is $R = \sum N_d / \sum E_d = \sum w_d r_d$, where $r_d = N_d / E_d$ and $w_d = E_d / \sum E_d$. It is an exposure-weighted average of daily rates. The unweighted daily mean and median answer different questions. A median may resist a corrupted day while also hiding a real high-volume incident; it does not reconstruct the period’s affected-unit risk.

Quantity [AI/PD]	Interpretation and boundary
Ratio of period sums	Rate for the compatible exposed population over the declared window; numerator/exposure capture must be aligned.
Median daily rate	Typical day under the chosen daily definition; not a substitute for exposure-weighted period risk.
Broad versus scoped rate	Descriptive scope comparison only if each numerator matches its own exposure and the windows are compatible.
Excluded-population rate	Subtraction requires verified nesting, common definitions and positive remaining exposure; labels alone do not establish these.

[HR/RC/AI] The register’s broader numerator and denominator carry different scope descriptions. If they are assumed to form one valid nested bridge, the arithmetic places greater risk outside the retained population, reversing the prose explanation. That conditional contradiction is diagnostic; without source reconciliation, neither group-risk story is established. A separate timeliness numerator belongs to another comparison and cannot be combined into this bridge.

[PD] Freeze a comparison manifest before recalculation: source versions, included days, denominator eligibility, recognition versus event date and excluded rows. For zero exposure, a rate is undefined; retain the exception and numerator for investigation. Do not discard genuine zero-defect days merely because they lack a positive numerator. Quarantine invalid records and quantify missing exposure rather than choosing an aggregation that makes the result plausible.

[AI] Multiplying numerator and exposure by the same positive factor leaves the rate unchanged. Correction factors from different metric families cannot decompose an observed discrepancy without a common replay that isolates the changed inputs.

Basis: registered scope and aggregation forks [HR]; ratio identities and conditional contradiction [RC/AI]; NIST weighted mean and measures of location.

TECHNICAL ANALYSIS

Arithmetic guards require conditions that can themselves fail

[AI] An invariant can reject an incompatible result without proving the accepted result correct. Its scope and assumptions belong in the release contract.

Proposed check [PD]	Necessary applicability conditions [AI]
Topline at least component sum	Components are disjoint, nonnegative subsets of the same topline numerator, with identical positive exposure, population, time and version.
Rate bounded by one per unit	Numerator counts at most one qualifying affected unit or opportunity per eligible denominator unit. Repeated-event intensity has no such automatic ceiling.
Period rate within daily range	Period uses exactly those compatible days and positive exposure weights. Passing the bound does not authenticate scope, values or maturity.
Identity collision rejection	Distinct intended constructs resolve to an unauthorized common binding; documented equivalent aliases require different treatment.

[AI] Individually contained buckets may overlap: their sum can exceed the parent even when each count is correct. An exhaustive and mutually exclusive partition gives equality; a disjoint partial partition gives only a lower bound. Negative adjustments, unmatched snapshots or a denominator attached to another dimension invalidate the simple inequality. Sibling metrics that exceed a parent cannot be presented as a reconciled additive explanation.

[HR/AI] The register records a low topline and an out-of-range period cell as rejected outputs. These are evidence of described analytical catches, subject to the premises above. They do not establish a measured false-release rate, continued software enforcement or the corrected value’s independent truth.

Use discriminating fixtures, not only known-good examples

[PD] Metamorphic tests should permute rows, split and recombine disjoint days, apply common scaling and introduce a documented alias without changing the intended result. Adversarial fixtures should change denominator scope, add overlapping components, alter a definition under the same name and preserve an obsolete cache. Specify the expected relation before execution. A deliberately changed construct must not be forced to remain invariant.

[AI] The original metamorphic-testing method motivates related follow-up cases when a complete output oracle is difficult. Its application here still needs independently justified semantic premises; passing relational checks cannot replace them.

Basis: registered guards and sibling limitations [HR]; conditional inequalities [AI]; Chen, Cheung and Yiu (1998). All new fixtures are proposed [PD].

TECHNICAL ANALYSIS

Routing accuracy must be assessed on meaning and context

[HR/AI] The register reports reduced keyword duplication and clean workbook checks, describing properties of an artifact. They do not establish that a request will select the right metric in context.

Measurement family	Validity and invariance audit [PD]
Identity and alias resolution	Independent concept signatures; include legitimate aliases, reused identifiers, similar labels and ambiguous requests.
Rate and burden	Audit numerator/exposure alignment, aggregation and maturity; compare like business lines and supported node populations.
Driver attribution	Verify that the native hierarchy or mapped sibling represents the requested mechanism. Count shares are not automatically rates or causal drivers.
Current binding and availability	Separate historical spelling, authenticated current definition, authorized read and timely valid output. Unknown is not zero.
Benchmark loss	Freeze semantic truth, permissible hold, deadline and nonoverlapping resource categories before assignment.

[AI] Shared family stems can be legitimate; a specific phrase may still route ambiguously despite having no exact duplicate. Conversely, repeated text may correctly express a shared concept. Audit context-to-concept decisions, not merely strings. Where several metrics legitimately apply, require a declared set-valued answer or clarification; forcing an arbitrary single choice changes the truth standard.

[PD] Sample a random core of 200 future metric-context cases from a declared frame. This is not 200 unique historical metrics. Two blinded coders independently judge identity, scope, applicable aggregation, currentness and release eligibility against a frozen codebook. Retain both labels and adjudicate using independent owner-supported evidence; unavailable truth remains unresolved.

[PD] Under independent cases, 200 observations give a worst-case raw-agreement margin of about ±6.9 percentage points for a normal 95% screen. Shared concepts/source families require a design-effect adjustment. Kappa precision requires prevalence-specific pilot simulation; report confusion matrices, class prevalence, raw agreement and intervals alongside kappa. Oversampled rare drift and alias exceptions need design weights for population estimates.

[AI] Agreement measures reproducibility, not correctness. Independent counterexamples must be able to overturn a widely shared definition. Blank-field and thin-definition discrepancies remain separate source versions until their coding rules are reconciled.

Basis: registered mapping, hierarchy and blank-field limitations [HR]; proposed blinded audit [PD]; Cohen (1960); W3C SKOS. No historical routing accuracy is estimated.

TECHNICAL ANALYSIS

Definition changes must not be mistaken for operating signals

[AI] A corrected display, a renamed metric and a newly filtered population can change a time series without changing customer experience. Process interpretation starts after comparability is established.

Separate chronological operating readings from artifact revisions

[PD] For ordered commensurate observations, $MR_t = |x_t - x_{(t-1)}|$; tentative individual limits are $\text{mean}(x) \pm 2.66 \times \text{mean}(MR)$, and the moving-range upper limit is $3.267 \times \text{mean}(MR)$. Retain order, spacing and phase. Short or degenerate sequences do not establish stable variation. An out-of-limit point prompts investigation; it is not an identified intervention effect.

[HR/RC/AI] The private inventory contains nine screens: three ordered input-metric quotation series with four observations each, and six mechanical version or unequal-window comparisons. All points and adjacent moving ranges lie inside their provisional limits. Incomplete date and identity crosswalks prevent a commensurate operating-panel claim. Different topline, weekly/daily readings and offline profile samples cannot be pooled. Raw coordinates and reconstructable pass/failure companions remain private.

[PD] A Laney p-prime chart additionally requires repeated compatible affected-unit counts and denominators, plus an interpretable subgroup/dispersion model. A large denominator or a rounded rate alone does not supply those inputs. Dispersion caused by changing definitions cannot be repaired by widening limits. Within-limit movements are not established improvement or deterioration.

Selection and context constrain interpretation

[HR/AI] The reported corrupt-site ranking is not a fixed worst-site cohort with an untreated follow-up. No historical regression-to-the-mean share is identified. As an explicit hypothetical screen, selecting the worst tenth of standard-normal errors gives a mean 1.755 SD; assumed repeat reliabilities 0.3, 0.6 and 0.9 imply untreated reversion of about 1.23, 0.70 and 0.18 SD. These are sensitivity assumptions, not measured project effects.

[PD] Freeze focus/non-focus membership and compare baseline level, trend, maturity and mix before intervention. Test business-line, node-type and source-family moderators with uncertainty. High exposure concentration makes median-versus-period disagreement more consequential; overlapping taxonomies weaken additive guards; unavailable definition owners reduce the usefulness of automated revalidation. Local success does not validate an unsupported network's metrics.

Basis: registered sequence and targeting limitations [HR]; proposed SPC and selection sensitivity [PD]; NIST individuals charts; Laney (2002); Barnett et al. (2005).

TECHNICAL ANALYSIS

Credit the measurement decisions without inventing implementation impact

[HR/AI] The register attributes metric-selection direction, diagnostic hypotheses and release rules to the project lead. Source-system ownership, app editing and live-query execution are distinct contributions.

Contribution claim	Process-tracing assessment [HR/AI]
Reframed the discrepancy and specified the metric contract	Provisional hoop: explicit derivative attribution supports direction; original authored instructions and their timing still require authentication.
Chose aggregation, unknown-state and offline boundaries	Provisional hoop for analytical design. The source’s population-risk explanation is corrected here rather than credited as a valid finding.
Executed queries or changed deployed code	Recorded execution also involves builder/analyst agents and separate code owners. Sole human execution is not established.
Extended a mapping artifact	The register describes delivery and cross-row review; artifact completeness does not demonstrate consumer adoption or semantic accuracy.
Improved outcomes or repaired the upstream estate	Unestablished: no causal operating comparison, mature benefit ledger or authenticated repair completion is supplied.

[AI] Repeated derivative summaries are not independent corroboration. A reported personal statement or session activity is not a smoking gun for the whole execution chain. No doubly decisive attribution test is established. The rejected hypotheses and deliberate non-change show analytical discipline only within the documented decision scope.

Consolidated limitations and reproducibility

[HR/AI] Evidence: the supplied register is derivative; original lookup payloads, code changes and build artifacts need authentication. Measurement: scope labels, denominator lineage, alias equivalence, row definitions, currentness and maturity remain material. Identification: neither before/after display agreement nor a guard catch identifies customer or cost impact. Transport: a constructed drift benchmark does not estimate live drift prevalence, authority availability or adoption.

[AI/PD] Economics uses hypothetical resource prices; reported corrected figures are not savings. Privacy keeps raw values, dates, identifiers, locators and inversion anchors private. Reproduction requires source/version manifests, exact signature comparisons, exception logs, test fixtures and executable calculations. The absence of a recovered completion trace does not establish whether further work occurred elsewhere; current source status requires fresh evidence.

Basis: role, governance and limitation records [HR]; process-tracing grades [AI]. Review remains conditional on original-source recovery and does not certify deployment.

TECHNICAL ANALYSIS

Evaluate whether bindings remain reliable when definitions change

[PD] Randomize independent analyst/source-dependency domains to MIRC or usual documentation in a controlled sequential benchmark. The domain, not an individual refresh, is the independent assignment unit.

[PD] Each domain begins with an authenticated binding and receives eight scheduled refreshes. An independent hidden truth manifest determines whether the definition changed before each deadline. A stale binding remains stale until a correct authorized rebind. Additional changes while stale do not make it valid. The same source-change and authority-availability law applies to both arms; no production permission or customer action is executed.

[PD] MIRC changes the review/revalidation policy. Valid aliases remain valid; redundant equivalent rebindings consume resources. When a change is detected, unavailable authority requires a hold rather than invented validation. A correct authorized rebind takes effect before that refresh's deadline. A missed stale binding can release the wrong result and retain the stale state.

[PD] The primary intention-to-treat estimand is $\Delta = E[L(\text{usual})] - E[L(\text{MIRC})]$, total prespecified resource-weighted loss over all eight original refreshes per assigned domain. Positive Δ is benefit. Loss distinguishes valid reuse, unnecessary or required rebind, hold and wrong release. Missing execution remains a hold when output is suppressed, or wrong release when a stale cached output persists. No completed-only or valid-only denominator is permitted.

[PD] Freeze business definitions, valid aliases, aggregation premises, response deadline and loss categories independently before randomization. Adjudicators are blinded to arm. Audit false releases, unnecessary holds, valid-output availability and authority compliance separately; an economic score cannot authorize unsupported output. Shared staff or dependencies must be clustered together, not relabeled as independent domains.

[PD] The finite state/reward planning model uses synthetic drift and response laws, exact boundary-specific power and same-horizon costs. Its sample size is conditional, not demonstrated feasibility or achieved field power. Pilot evidence must estimate drift, reviewer dependence, costs and authority availability before recruitment or deployment.

[PD] A subsequent operating rollout could stagger adoption across independent dependency clusters. Estimate group-time ATT against not-yet-treated units with no anticipation, common support, baseline/event-time and equivalence diagnostics, and interference controls. Nonsignificant pretrends do not prove comparability. Freeze mature customer eligibility and obtain counts/covariance for a separately powered field design; this benchmark cannot establish customer benefit.

Basis: proposed sequential semantic-binding benchmark [PD]; synthetic model pages provide probability, cost and power assumptions. Later field design follows Callaway–Sant'Anna; it remains unpowered.

TECHNICAL ANALYSIS

Four propositions separate reliable reference from reliable performance

[AI/PD] MIRC connects semantic identity, aggregation premises, current evidence and accountable release. Its average benefit is a testable proposition, not a result established by writing the standard.

Proposition and mechanism [PD]	Disconfirming result or boundary [PD]
P1: a versioned signature reduces distinct-construct misbinding while preserving legitimate aliases.	Fresh independent tasks show no material reduction or more erroneous alias holds; unavailable authority limits the mechanism.
P2: matched population and time contracts reduce wrong period-rate releases.	Releases remain wrong after an authenticated replay, or daily medians are still presented as exposure-weighted risks.
P3: conditional guards improve detection without excessive valid-output suppression.	Overlap, repeated-event or equivalent-alias fixtures cause more false holds than the preregistered tolerance.
P4: dated cache authority and accepted-owner routing reduce stale-value release after definition changes.	Wrong releases recur despite detected change, or useful valid output deteriorates beyond its guardrail.

[AI] SKOS separates lexical similarity from conceptual mapping. NIST location and weighting definitions show why robust daily summaries need a different label from period exposure rates. Metamorphic testing supplies fault-revealing relations; MIRC adds an explicit applicability contract so a correct non-additive metric is not condemned by a false invariant. This is a proposed synthesis, not a claim to have invented these methods.

[AI] Deming’s tampering warning is relevant only after measurement is meaningful; MIRC checks that prerequisite. Campbell and Goodhart motivate testing whether target pressure changes classification or selective omission, without alleging manipulation here. Ittner and colleagues distinguish measurement sophistication from performance. Tucker and Edmondson distinguish workarounds from organizational learning: an offline reference can sustain a narrow task while the upstream defect remains unresolved.

[AI/PD] Composite-indicator design requires explicit weighting and compensation. Identity, authority and maturity are mandatory gates, not interchangeable points in a quality score. Do not average separately rebased groups into a network rate; restore compatible quantities and exposure weights. A high completeness score cannot repair a wrong metric.

[PD] Preregister material effect and guardrail thresholds. A lower confidence bound above economic break-even supports the modeled investment gate; an upper bound below it contradicts that economic claim. An interval spanning the threshold is inconclusive. Statistical success cannot override an unresolved mandatory source condition, and benchmark success cannot be relabeled customer impact.

Basis: proposed MIRC propositions [PD]; W3C SKOS; NIST; Chen et al.; Deming; Campbell; Goodhart; Ittner et al.; Tucker and Edmondson; OECD. No external effects imported.

PROSPECTIVE MODEL / BINDING RELIABILITY

Test whether a correct binding stays correct as its meaning changes

[PD] Randomize independent metric-dependency domains to the Metric Identity and Release Contract (MIRC) or usual pinned-ID reuse. Every value here is a hypothetical planning assumption.

[PD] Authenticate each domain at time zero. Follow eight scheduled refresh intervals. Before each deadline, an unobserved catalogue change can make the cached binding stale; additional changes leave it stale. The hidden reference fixes concept, population, denominator, transform and authorized version. A valid alias remains valid.

State and action at the deadline	Resource loss	Next cache state
Valid reuse	1	Valid
Authorized correct rebind	3	Valid
Hold, including missing output	6	Unchanged
Wrong-semantic value released	12	Stale

[PD] Primary ITT is usual-minus-MIRC total loss across all eight original deadlines, averaged over assigned domains. Detected staleness with unavailable authority must remain a hold; it is not a repair. Missed staleness produces a wrong release and persists. A correct authorized rebind repairs the same-deadline output and subsequent cache.

DO NOT REWARD SILENCE

[RC/PD] All-hold loss is 48 per domain; usual-policy loss is 30.12 and full-protocol loss is 18.76. Legitimate permission waiting still has resource burden. No customer harm or historical savings are assigned.

[PD] Controlled replay of binding lifecycles; no real source query, grant or publication. Independent blinded semantic adjudication must assess the reference itself and record disagreements.

PROSPECTIVE MODEL / BINDING RELIABILITY

A stable identifier is not a permanent validity certificate

[PD] Shared change exposure makes the refreshes dependent. Eight observations inside one domain do not create eight randomization units.

Planning primitive	Usual policy	Full MIRC
Stale-binding detection	0.45	0.80
Authority available after detection	0.75	0.75
Valid: reuse / redundant rebind / hold	0.90 / 0.06 / 0.04	0.94 / 0.03 / 0.03

[PD] Equally likely domains have per-interval change probabilities 0.10 or 0.30; events are independent only conditional on type. The protocol cannot alter the scripted source truth or approval authority. Increasing policy parameter lambda linearly changes detection and valid-state action probabilities; repair changes later state occupancy and variance.

[PD] Compute a two-state, accumulated-loss probability distribution at each refresh, then mix domain types. Convolve independent arms. A shared-uniform coupling orders costs downward and valid-cache states upward with lambda; common drift preserves this order. The economic boundary is therefore least favorable within this known family.

Required field calibration	Condition before interpreting results
Reference and missingness	Dual-code semantic validity blind to arm; retain withdrawals and timeouts. Missing execution is a hold if suppressed, or a wrong release if stale output persists.
Guardrails and transport	Separately size wrong release, unauthorized rebind and usable-output coverage margins. Unknown production drift, misrepair and dependency can invalidate this planning law.

A SCOPED TEST

[PD] Perfect authorized repair is a benchmark assumption. It is not evidence that real repairs are correct. Freeze policy versions and isolate shared maintainers, caches and authority dependencies before assignment.

[PD] Within-domain Markov dependence; independent domains and arms. The design does not isolate the effect of a single ID rule or establish operational adoption.

PROSPECTIVE MODEL / BINDING RELIABILITY

Size the assurance test against the cost it must cover

[PD] Require 80% power to detect break-even against zero and 80% probability that a one-sided 97.5% lower bound exceeds break-even at the full planning alternative.

[RC/PD] Usual eight-refresh loss is indexed to 100. The full planning effect is 37.72 points. Net resource value = reduction × avoidable share a × non-overlap h – incremental cost 10. Base a = h = 1 gives break-even 10 points.

Domains per arm / total	80% MDE	Economic power	Both gates
80 / 160	28.19	83.3%	Fail
673 / 1346	10.00	>99.9%	Conditional
840 / 1680	8.96	>99.9%	Conditional

[RC/PD] Base minimum: 673 per arm. Drift heterogeneity 0.05 / 0.35 requires 726. Candidate 840 per arm clears that stress. This expanded-portfolio plan exceeds the historical evidence base; recruitment feasibility is unestablished. Calibrate dependence and resource weights before recruitment.

Net at full effect / cost	a = 0.50; h = 0.80	a = 0.75; h = 0.90	a = h = 1.00
7	8.09	18.46	30.72
10	5.09	15.46	27.72
13	2.09	12.46	24.72

[PD] Cost 10 allocates setup 4, metadata maintenance 3 and production verification 3 to the same fixed service pool. Deadline handling is already in loss; never charge a shared incident twice. Research adjudication is separate. Shares scale net burden uniformly; otherwise reprice categories.

[PD] Future field value = eligible treated volume × mature customer ATT reduction × avoidable cost/outcome × non-overlap – incremental cost. Population and maturity must match. Missing inputs leave this separate claim unestimated and unpowered.

[PD/RC] Exact known-law two-sided alpha ≤ 0.05; economic false clearance ≤ 2.5%. Nuisance uncertainty, false repair and guardrail power require calibration; conditional precision is not field assurance.

PROSPECTIVE MODEL / BINDING RELIABILITY

Buy evidence before deciding whether to retain, rebind or route

[PD] Information value addresses a separate planning decision. Better evidence can redirect effort without authorizing a publication or producing a customer benefit.

[PD] States are authorized-current, authorized-changed, and unresolved authority, with priors 0.45 / 0.35 / 0.20. Unresolved authority takes precedence. The remaining states partition current versus changed meaning. Other failures require an expanded state space.

Planning action / state	Current	Changed	Unresolved
Prepare retained binding	10	-20	-28
Prepare authorized rebind	3	16	-16
Route and hold	-4	-3	8

[RC/PD] Prior-optimal rebind planning is worth 3.75; perfect information adds at most 7.95. Three-label signal accuracy is 0.50, 0.75 or 0.90; each incorrect label receives half the remaining probability. Select the best plan after observing the signal.

Signal quality	EVSI	Evidence cost	Net EVSI
Low	0.0000	0.50	-0.5000
Base	2.8625	1.50	1.3625
High	5.9150	3.00	2.9150

[PD] Utilities already include subsequent action costs. Evidence acquisition is additional. These utilities are separate from, and must not be added to, trial resource value. No probabilistic recommendation waives identity, population, denominator, transform, version or release-authority requirements.

THE MANAGEMENT DECISION

[AI/PD] Fund an independently verified binding lifecycle, with explicit retain, rebind and hold states. An exact identifier supports reproducibility; current semantic evidence determines whether its output is fit for release.

[PD/RC] Hypothetical finite-state Bayesian decision model. Historical prevention, adoption, financial savings and mature customer impact remain unestablished.

Reproduce the reporting logic while protecting the operating evidence

[AI] This is a measurement-governance and operating-design case. The source describes corrections and standards; the prospective analysis asks whether a formalized binding policy would improve repeated release decisions.

Claim class	Meaning and boundary
[HR] Historical record	Registered observation, correction, artifact description or attributed decision. Underlying source artifacts and recorded descriptions remain distinct evidence levels.
[RC] Recalculated	Protected display comparisons, source reconciliations, synthetic examples and planning-model outputs recomputed from declared inputs. Arithmetic does not authenticate provenance.
[AI] Analytical inference	Conditional scope contradiction, aggregation and invariant limits, or bounded interpretation. A logically valid conclusion still depends on its stated premises.
[PD] Proposed design	MIRC, future tests, source-owner responsibilities, economics and information-value scenarios. These are not executed experiments or realized benefits.

Reproduction sequence

[PD] Verify the immutable workbook and frozen extract. Recompute each reported value using the exact numerator, denominator, scope and precision recorded. Preserve incompatible versions, undefined periods and unresolved lineage. Test every invariant on both valid and deliberately incompatible fixtures.

[PD] Recreate the Markov reward law, independent-arm tests, economic sensitivities and finite-state information model from their declared assumptions. The private package contains raw evidence, exact cell locators, source searches, code, review dispositions and final file hashes. It excludes production credentials and the full original workbook.

WHAT THE NEW ANALYSIS CHANGES

[AI] The source’s scope explanation is not supported by its own conditional arithmetic. A robust daily median also changes the period-risk question. These findings require a semantic contract and unresolved states, not a more confident dashboard.

Labels: [HR] historical record; [RC] recalculated; [AI] analytical inference; [PD] proposed design. Source basis S1; see private annex.

RESEARCH REFERENCES / METHODS AND INTERPRETATION

Define identity, aggregation and currentness before release

Primary research and official methods inform the analysis and proposed test. They do not independently corroborate the historical project.

Miles, A., & Bechhofer, S., eds. (2009). SKOS Simple Knowledge Organization System Reference. W3C Recommendation. Concepts, labels and mappings are distinct. The case adapts that distinction for approved metric aliases; SKOS does not validate local semantic equivalence.

W3C (2024). Data Catalog Vocabulary (DCAT)—Version 3. Official Recommendation. Catalog and version metadata support traceable resource descriptions. A version record is not a certificate of accuracy or current business fitness.

NIST (1997). DATAPLOT Reference Manual: Weighted Mean, p. 2-65. Official formula reference. The case derives the ratio-of-sums and convex-hull constraints using positive eligible-exposure weights; no local effect is imported.

NIST/SEMATECH. e-Handbook of Statistical Methods, §1.3.5.1: Measures of Location. Mean and median describe different location functionals. Robustness to extreme values does not turn a daily median into an exposure-weighted period risk.

Chen, T. Y., Cheung, S. C., & Yiu, S. M. (1998). Metamorphic Testing: A New Approach for Generating Next Test Cases. HKUST-CS98-01; author-posted 2020. Related follow-up tests can reveal missed faults. Passing a relation still does not establish semantic truth.

Fielding, R., Nottingham, M., & Reschke, J., eds. (2022). HTTP Caching. RFC 9111. Official freshness and validation semantics provide a design analogy. The local profile reference is not asserted to be an HTTP cache; historical availability does not establish current origin validation.

Moreau, L., & Missier, P., eds. (2013). PROV-DM: The PROV Data Model. W3C Recommendation. Versioned entities, activities and derivations inform the source ledger. Provenance represents a derivation claim; it does not prove correctness.

Research access scope is recorded in the private memo, distinguishing fresh verification, prior-series verification, publisher metadata and full-text review.

RESEARCH REFERENCES / METHODS AND INTERPRETATION

Separate diagnostic evidence from causal and statistical conclusions

Primary research and official methods inform the analysis and proposed test. They do not independently corroborate the historical project.

NIST/SEMATECH. e-Handbook of Statistical Methods, §6.3.2.2: Individuals Control Charts. Moving ranges require ordered observations and an interpretable process phase. Short registered sequences receive tentative screens, not claims of stable operational improvement.

Laney, D. B. (2002). Improved Control Charts for Attributes. *Quality Engineering*, 14(4), 531–537. Publisher abstract. Attribute-chart dispersion requires a defensible model; rounded rates and unknown or changing exposure do not support a valid large-sample rate chart.

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20, 37–46. Publisher metadata. Kappa supplements the proposed raw agreement, prevalence and confusion matrix; agreement cannot authenticate the reference labels.

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. *IJE*, 34, 215–220. Primary abstract. Selected extreme units can recover without intervention; the case's numerical sensitivity is explicitly hypothetical.

Collier, D. (2011). Understanding Process Tracing. *PS: Political Science & Politics*, 44, 823–830. Primary publication. Evidence tests distinguish direction, authorship and execution. A derivative attribution is not a smoking gun for deployment or outcome ownership.

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 the optional later field design; this is not the randomized Markov-benchmark estimator.

Wang, R. Y., & Strong, D. M. (1996). Beyond Accuracy: What Data Quality Means to Data Consumers. *JMIS*, 12(4), 5–33. Original article scan. Fitness for a defined decision requires more than an accessible identifier, a populated field or a plausible number.

Research access scope is recorded in the private memo, distinguishing fresh verification, prior-series verification, publisher metadata and full-text review.

RESEARCH REFERENCES / METHODS AND INTERPRETATION

Connect reporting discipline to learning and investment decisions

Primary research and official methods inform the analysis and proposed test. They do not independently corroborate the historical project.

The W. Edwards Deming Institute. The Funnel Experiment. Official exposition. Responding to routine variation can worsen a process. MIRC adds the prior requirement that metric definitions and populations remain comparable; no historical tampering effect is claimed.

Campbell, D. T. (1976). Assessing the Impact of Planned Social Change. Original institutional record. Consequential indicators can distort underlying behavior. A clean audit or high completion score should not replace independently valid release evidence.

Goodhart, C. A. E. Problems of Monetary Management: The U.K. Experience. 1975 conference; revised volume listed as 1976 by the RBA. Publisher archive establishes provenance of Goodhart's law. The case tests risks of targeting proxies without alleging historical manipulation.

Ittner, C. D., Larcker, D. F., & Randall, T. (2003). Performance Implications of Strategic Performance Measurement in Financial Services Firms. AOS, 28, 715–741. Author manuscript. This case does not infer operating performance merely from a more sophisticated measurement architecture.

Tucker, A. L., & Edmondson, A. C. (2003). Why Hospitals Don't Learn from Failures. CMR, 45(2), 55–72. Author working paper. A local workaround differs from source-level organizational learning; a spelling cache does not establish repaired upstream governance.

OECD / European Commission JRC (2008). Handbook on Constructing Composite Indicators. Official handbook. Weighting, normalization and compensation require explicit choices. Mandatory source conditions cannot be offset by a high aggregate quality score.

Strong, M., Oakley, J. E., Brennan, A., & Breeze, P. (2015). Estimating the Expected Value of Sample Information Using the Probabilistic Sensitivity Analysis Sample. MDM, 35, 570–583. Author repository. The case uses a separately derived finite-state decision model, not this article's regression estimator.

Research access scope is recorded in the private memo, distinguishing fresh verification, prior-series verification, publisher metadata and full-text review.

FINAL REFEREE REPORT / INTERNAL AGENT-ASSISTED REVIEW

Approve the reporting case within its verified evidence boundary

[AI] This is an author-led review supported by separate evidence, methods and numerical agents. It is not external peer review, client approval or certification of a firm ranking.

Major issue	Disposition in the revised case
Scope and rate arithmetic	The source’s risk-direction story is rejected conditionally; incompatible geography labels prevent an observed subgroup ordering. Display corrections remain separate from operating effects.
Identity and guard validity	Legitimate aliases, disjointness, numerator bounds and positive matched exposures are explicit. Passing a conditional guard does not certify the source.
Aggregation and currentness	A daily median has its own estimand. Historic spelling support cannot validate current values. Source precision, workbook counts and incomplete dispositions remain documented forks.
Causality and economics	The dynamic benchmark includes every deadline. The initial design cannot detect break-even; a larger stress-tested candidate remains conditional on recruitment and calibration. Information value stays separate from field impact.
Evidence and reproduction	Personal direction is graded separately from execution and platform ownership. Raw evidence is private; arithmetic and final rendered pages are checked and bound to file hashes.

Partner critique: would a COO act on page one?

[AI] Yes: require a verified definition and population for every published KPI, hold unsupported subgroup comparisons, and assign source owners to changes that reporting teams cannot resolve. The cover names the decision, a new finding and the evidence boundary.

Minor revisions and remaining limitations

[AI] Review preserves approximate displays rather than claiming exact equality, narrows universal numeric guards and distinguishes mapping completeness from semantic accuracy. Source recovery and historical outcome limits remain explicit; the proposed benchmark requires independent calibration and feasibility checks.

REVIEW DISPOSITION — PASS WITHIN STATED SCOPE

[AI] The case supports documented measurement decisions, new reconciliation findings and a reproducible proposed evaluation. It establishes neither sustained enterprise adoption nor historical customer improvement or realized financial benefit.

Labels: [HR] historical record; [RC] recalculated; [AI] analytical inference; [PD] proposed design. Source basis S1; see private annex.