

OPERATING CASE STUDY / OPERATIONAL EXCELLENCE

Execution Assurance & Accountability

From assigned work to verifiable execution

An anonymized project case study of a distributed service network's action-governance redesign: how evidence standards, execution maturity and verification capacity determine whether corrective work can be trusted.

PROJECT CONTEXT / DISTRIBUTED SERVICE OPERATIONS

BLUF / BOTTOM LINE UP FRONT

Define the unit of verification.
Prove the gate before scaling the portfolio.

RECALCULATED FINDING

A regional action limit and a local assignment limit can imply opposite workload changes. The missing link is a defined verification unit—not evidence that a smaller action count reduced work.

CONTRIBUTION IN THE SOURCE RECORD

Advisory design participation and audit-data support.
Registry-attested role.

EVIDENCE STATUS

Operating-design case. Adoption, customer impact and savings remain unestablished.

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

EXECUTIVE ABSTRACT

Approve verifiable actions before expanding the action portfolio

An operating case on the design of execution assurance in a distributed service network. The record establishes a governance redesign; it does not establish its customer or financial effect.

CONTEXT | A mandate to act exceeded the evidence that action occurred

[HR] A network quality mechanism assigned corrective work across regional and local operations. The redesign proposed central action development, a mandatory validation gate, execution-maturity tiers and recurring specialist verification. Later records describe tier alignment and local readiness audits. A later central reporting extract was empty, while a live execution-score panel was not established in the supplied register.

OBJECTIVE | Make action assignment, execution and effectiveness separately auditable

[PD] Convert each approved action into an evidence-carrying contract: a defined eligible opportunity, accountable operator, independently checkable execution event, capacity allowance and mature outcome window. Treat missing evidence as unresolved status; preserve the ability to observe work completed outside the central log.

APPROACH | Reconcile provenance before quantifying performance

[AI] This case audits the project register, separates advisory proposals from recorded implementation and formalizes the verification workload. [PD] It specifies an independent shadow test of gate decisions, a subsequent policy evaluation and a normalized decision-economics model. These are reproducible designs, not reconstructions of an unavailable trial.

FINDINGS | The action cap changes meaning when its denominator changes

[AI] A regional action-template limit and a local action-assignment limit do not define the same workload. Depending on whether regional templates require local verification, the apparent change can increase or reduce check demand. The source does not resolve this mapping. A blank reporting extract also cannot identify a zero-execution baseline.

DECISION VALUE | Validate the unit, the gate and the capacity before scaling

[PD] Authorize a bounded evidence-validation pilot. Require a reconciled action-instance ledger and measured verification minutes before expanding the portfolio. Model the value of correcting unsupported decisions separately from customer impact; count neither as realized benefit before observation.

IMPLICATIONS | Credit the contribution the record supports

[HR] An assistant-maintained, advisor-associated working-session record contains intake, capacity, escalation and retirement proposals; another author originated the overhaul. A later registry entry assigns the advisor a specific audit-data sharing action. [AI] The supported portfolio claim is registry-attested design-session association and a narrow assignment; human origination, implementation influence and impact remain unproven.

S1: master, narrative and linked evidence E01–E06, E17; selection and claim grades in the private ledger. Labels: HR historical record; RC recalculated; AI analytical inference; PD proposed design.

EXHIBIT 1 / EVIDENCE STATES

An empty reporting view does not prove that no work occurred

[AI] The central visibility gap is a valid finding. A zero-execution baseline is not.

FOUR DISTINCT EVIDENCE STATES

Historical record at register level • statements have different scope



An incomplete reporting link is compatible with some local execution.

Approval is not implementation. Recorded completion is not independent proof.

No time axis or measured adoption rate is implied by this evidence map.

[HR] The register describes a written validation requirement, subsequent tier alignment and local readiness-audit activity. It also records an empty programmatic action block in a later reporting artifact and no observed execution-score values. These observations concern different evidence surfaces. They can coexist without proving either full adoption or complete non-execution.

Decision error	Why it fails	Required correction
Call the baseline zero	The expected action-opportunity set and capture rate are absent.	Recover the eligible ledger; reconcile central and local records.
Call the rollout complete	Framework alignment is not a scored, operating verification system.	Separate approval, configuration, use and validated execution.
Call the redesign ineffective	No mature treated-versus-counterfactual outcome comparison exists.	Estimate policy effects prospectively after establishing measurement.

EXECUTIVE DECISION

[PD] Keep missing, failed, not due and not applicable as different states. Do not turn a reporting defect into an operating-performance verdict.

[HR] S1:E01, E04–E07, E13–E15; M07–M08. [AI] Register-level evidence adjudication. [PD] Missingness and capture protocol; Rubin (1976), reference R3.

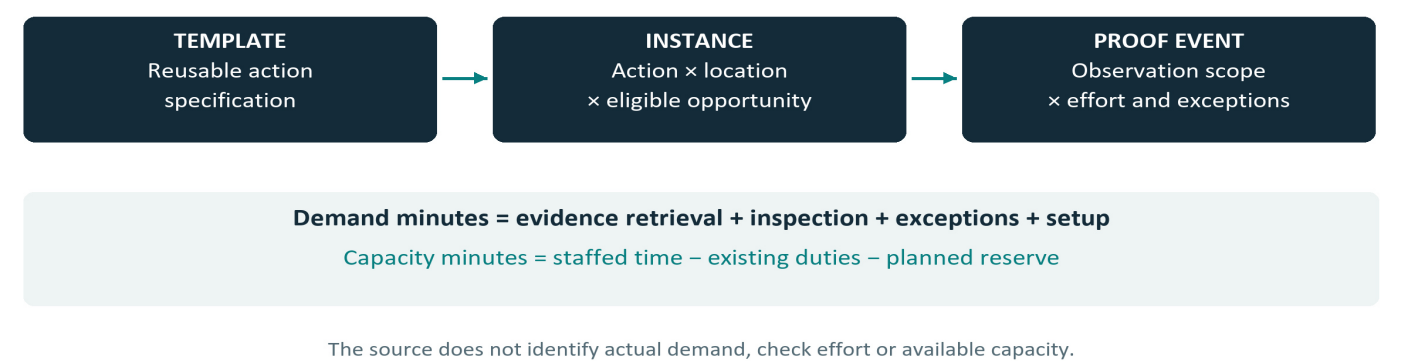
EXHIBIT 2 / ORIGINAL ANALYTICAL FINDING

The action cap is undefined until the verification unit is fixed

[AI] A smaller action count can create more verification work. A larger count can reuse the same evidence. The source caps alone cannot decide which occurred.

COUNT THE WORK THAT VERIFICATION REQUIRES

Proposed unit contract • conditional implications of the recorded caps



[RC] The private annex evaluates alternative unit mappings. A regional summary check expanded into local assignments increases check demand in one interpretation. If the regional card already requires separate local checks, the local cap can reduce demand. A separate staffing assignment changes the comparison again. No branch establishes actual labor hours.

Contract term	What must be fixed before approval
Instance	One action type × location × eligible operating opportunity; not a template title.
Verification event	Which instances one observation validly covers; required recency and sampling frame.
Effort	Observed retrieval, inspection, exception handling and adjudication minutes per event.
Capacity	Specialist minutes available after existing operational duties, including demand variability.

[AI] The finding changes the decision from “choose the tighter cap” to “define the unit and fund its verification.” It does not establish that the redesign overloaded specialists or caused the missing reporting record.

RELEASE RULE

[PD] Approve additional assignments only when their marginal verification demand fits a measured capacity envelope and the evidence remains valid at the proposed coverage.

[RC] S1:M06–M07, E03–E04; private/capacity_audit.py and capacity_audit.json. Public ratios withheld to prevent reconstructing the internal footprint. Little (1961), R2.

EXHIBIT 3 / EXECUTION MATURITY

Readiness, execution quality and sustained performance need different proof

[HR] Recorded tiers progress from prerequisites to adaptive execution and reduced oversight. [PD] The four proof states below are a proposed assurance design, distinct from the recorded tiers.

Stage	Proof to seek	Invalid shortcut
Readiness	A prerequisite is available in the relevant shift, location and operating condition.	A photograph or installed tool proves routine use.
Execution	The operator performed the specified behavior in an eligible opportunity; an independent artifact supports the event.	A task was marked complete, so the work occurred.
Adaptive quality	The operator recognized an exception and chose the prescribed response within the applicable window.	A binary presence check establishes judgment quality.
Sustainment	Coverage and quality persist across shifts, demand states and reduced observation, with a re-entry trigger.	A short run of favorable scores justifies permanent graduation.

[AI] Maturity is a property of the operating behavior, not the difficulty of the assignment. A simple action may remain unreliable under a different shift or demand mix; a technically complex action may already be routine. Combining these dimensions in one tier obscures the intervention needed.

[PD] Retain distinct fields for readiness, delivery fidelity, operator competence, eligible exposure and outcome maturity. A failed prerequisite routes to the resource owner. An execution omission routes to the operating owner. A recurring system barrier routes to the program owner. An unobservable event routes to measurement repair.

[HR] The advisory record includes numerical retirement and escalation proposals, but later adoption of those thresholds is not established. [PD] Set release thresholds after estimating error, persistence and required coverage; document the rule before inspecting results.

WHAT THIS ENABLES

[PD] Standardize the control when execution survives reduced oversight. Reopen it when independent sampling detects drift, denominator change or a new operating context.

[HR] S1:E01, E03–E04, E06–E07, E13–E15; M15. [AI/PD] Case synthesis using implementation-fidelity dimensions, Carroll et al. (2007), R1.

EXHIBIT 4 / OPERATING DESIGN

Every approved action should carry its own evidence contract

[PD] The Evidence-Carrying Action Contract turns a validation principle into an assignable, testable unit of work. It is a proposed Luna Sol case framework, not a claimed deployed product.

Contract question	Minimum recorded field	Failure route
What changes?	Action specification, mechanism, operator, location, eligible opportunity and exclusions.	Return to designer when the behavior cannot be distinguished from the desired outcome.
What proves execution?	Timestamped artifact, evidence owner, independent check, taxonomy and retention rule.	Hold assignment if evidence proves only system state or task closure.
Can it be verified?	Instance count, coverage, observed effort, capacity owner and exception allowance.	Reduce scope or add capacity; record the tradeoff.
When can it be judged?	Action event, response hypothesis, outcome event date, maturity rule and data-freeze version.	Hold the effect decision; continue execution verification.
Who resolves failure?	Operating owner, systemic barrier owner, escalation trigger, retirement and re-entry rules.	Route the cause to the owner who can change it.

[HR] The advisory record identifies an alert that can disappear when a workload condition changes. [AI] Its disappearance cannot prove that the assigned person performed the requested response. Using clearance alone as proof would create a measurement-design failure even when the operational condition improves.

[PD] Use a replayable event trail: eligibility → assigned instance → action performed → evidence captured → independently adjudicated → outcome mature. Preserve each timestamp and correction. A verification hold must not silently authorize execution, and it must not erase the assignment from the denominator.

CONTROL OBJECTIVE

[PD] Produce fewer unsupported decisions per eligible opportunity without increasing missed execution, unsafe delay or hidden verification labor.

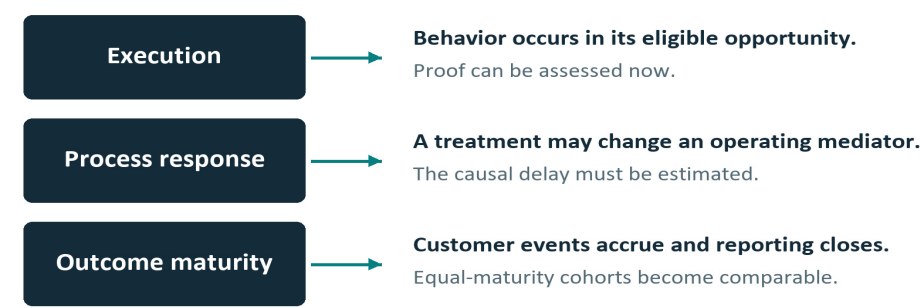
[HR] S1:E01, E03, D01–D08. [AI/PD] Proposed contract and routing. Implementation is not established by conceptual similarity to later tier definitions.

EXHIBIT 5 / DECISION TIMING

Separate the execution clock from the outcome clock

[AI] A completed action, a predictive lead and a mature customer outcome answer different questions. A single review date cannot make them interchangeable.

ONE ACTION • THREE DIFFERENT CLOCKS



Review eligibility depends on event time and a declared maturity rule, not the meeting date.

[HR] The register cites adjacent analyses of predictor lags and outcome reporting delays. These do not estimate this redesign’s impact. [AI] Contemporaneous association and temporal prediction answer different questions; their descriptions are not automatically contradictory.

Clock	Eligible decision
Execution	Did the prescribed behavior occur in its eligible window? Decide from action-event evidence.
Process response	Could the action plausibly have changed the mediator yet? Estimate this after specifying treatment; a predictor lag is not a causal response time.
Outcome maturity	Have affected customer events accrued and been reported completely? Freeze cohorts by event time and declared maturity.

[PD] Specify total observability from the action event to the mature outcome under one clock definition. Do not add lags that already include each other. Track reporting revisions separately from operating changes, and compare equal-maturity cohorts.

REVIEW DISCIPLINE

[PD] Continue checking execution while outcome evidence matures. Do not switch actions merely because the first review is too early to observe their effect.

[HR] S1:E02, E03, E11; M02, M05, M12–M13. [AI/PD] Distinct clock and estimand interpretation. No historical lag values or adjacent model effects are published as redesign results.

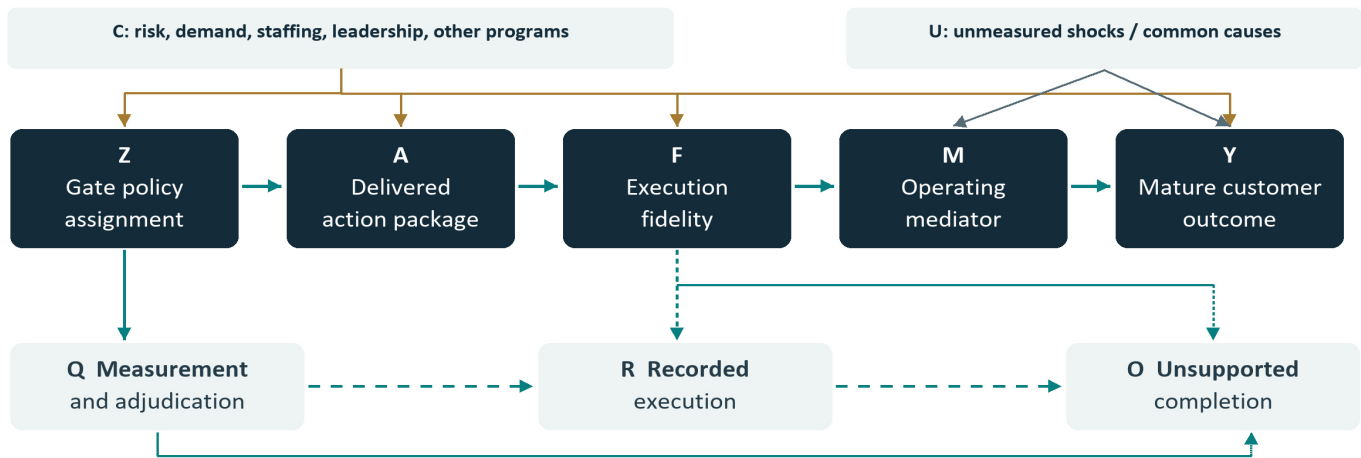
EXHIBIT 6 / THEORY OF CHANGE

Better records can improve measured performance without improving service

[PD] Evaluate the governance policy and the physical action package separately. The diagram makes the recording pathway explicit.

PHYSICAL CHANGE AND RECORDING CHANGE ARE SEPARATE PATHS

Proposed DAG • all effects are hypotheses • no historical effect is estimated



C also influences Q; dotted links are measurement/adjudication dependencies. O requires independent truth.

[PD] Z is assignment to the verification-gate policy; A is the delivered action package; F is execution fidelity; M is an operating mediator; Y is the mature customer outcome. Q is the recording and adjudication regime, R is recorded execution and O is the unsupported-completion endpoint. Dotted links carry true execution and measurement into observed decisions.

[AI] A rise in recorded execution could follow $Z \rightarrow Q \rightarrow R$ with no change in F or Y. A lower unsupported-completion rate could also follow blanket refusal to close work. This is why the proposed primary denominator is the independently enumerated eligible opportunity set and why missed valid completions, verification delay and customer outcomes are guardrails.

[PD] C names baseline quality, demand and product mix, staffing, local leadership and concurrent programs; these can affect assignment, fidelity and outcomes. U represents unmeasured shocks and behavior. The technical layer assigns the effect questions to this graph and identifies where the available record cannot close a causal path.

ATTRIBUTION BOUNDARY

[AI] No customer or cash improvement is attributed to this redesign. The record supports advisor-associated proposals, a narrow audit-data assignment and an evaluable mechanism.

[PD] Proposed DAG informed by S1 and Hernán & Robins, R5; Rubin, R3. Solid causal edges are hypotheses, not estimated effects. Reduced graph; measurement and confounding details follow.

EXHIBIT 7 / DECISION CHARTER

Fund a bounded verification pilot with explicit release and stop rules

[PD] The next investment should resolve whether the gate improves decisions at an acceptable operating cost. Customer-value scaling needs its own mature-outcome test.

Gate	Required evidence	Decision
Measurement ready	Reconciled eligible instances; replayable artifacts; independently adjudicated labels; reliable taxonomy.	Start a shadow assessment. Keep the current operating authority intact.
Decision quality	Paired decisions on the same frozen opportunities; false authorizations and false holds; lower error at a funded verification cost.	Advance only if decision economics and service guardrails support a live policy test.
Policy works	Randomized rollout by non-overlapping operating teams; complete eligible denominator; interference and workload monitored.	Estimate the intention-to-treat policy effect without removing rejected actions.
Customer value	Equal-maturity customer cohorts; action delivery recorded; avoidable cost and overlap validated by finance.	Scale the operating program only after its customer-value threshold is testable.

[PD] The numerical model later in this document is a synthetic decision-support example. It compares the detectability of an error reduction with the break-even for that same decision error. It does not substitute a gate-accuracy result for customer savings.

[PD] The executive sponsor owns release; the operating lead owns action delivery; an independent measurement lead owns adjudication; finance owns cost boundaries; the capacity owner approves verification effort. Publication retains design, exposure, execution and outcome as separate evidence states.

PAGE-ONE COMMITMENT

[PD] Approve a reversible learning stage, not an unverified rollout benefit. The first deliverable is a reconciled action ledger with a measured verification workload and a preregistered decision test.

[PD] Evaluation and decision charter; Carroll et al., R1; Rubin, R3; Callaway & Sant’Anna, R8; Strong et al., R12. Source record contains no completed evaluation or finance-validated savings.

The gate can change what is recorded before it changes what is executed

[PD] Separate governance assignment, actual behavior, verification and mature customer outcomes.

[PD] Let Z denote policy assignment; A the action package; F execution; M the operating mediator; Y mature customer harm; Q the measurement regime; R recorded execution; and O unsupported completion. Hypothesized paths are $Z \rightarrow A \rightarrow F \rightarrow M \rightarrow Y$ and $Z \rightarrow Q \rightarrow R$, with $F \rightarrow R$. O compares R with independently adjudicated F under Q. Confounders C and unmeasured causes U may affect assignment, fidelity and outcomes.

Quantity	Definition and identification boundary
Customer policy effect	[PD] $ITT(H) = E[Y_i(1,H) - Y_i(0,H) \mid \text{eligible}]$. Compare assigned protocol versus current practice for all preregistered eligible clusters over horizon H; retain noncompliers.
Unsupported closure	[PD] Difference in independently adjudicated unsupported-completion incidence over a common eligible action-opportunity set. Do not condition the denominator on admission or claimed completion.
Delivered-action effect	[PD] $E[Y_i(a^*,H) - Y_i(a_o,H) \mid \text{eligible}]$ for a specified action package versus current practice. This is distinct from adopting the governance protocol or selecting compliant sites.

[PD] Fix eligibility, package version, comparison, response lag and outcome-closure lag before assignment. Require a shipment to mature before inclusion in Y. Keep program exposure and execution separate; conditioning on R or voluntary adherence can select units and block causal paths.

[AI] Historical assignment, execution and mature-outcome panels are unavailable here. The graph formalizes hypotheses; it does not identify their effects. Residuals could reflect omitted paths, measurement error or noise, not a uniquely diagnosed cause.

Source S1: registry design and implementation records. DAG and estimands are proposed; raw evidence crosswalk remains private.

TECHNICAL FOUNDATION • MEASUREMENT VALIDITY

A complete-looking action score can conceal missing evidence and changing eligibility

[AI] Verification requires an explicit action unit, opportunity denominator and independent evidence rule.

Construct	Required definition and validity check
Action and opportunity	[PD] Separate a reusable template, a site-action instance and an execution opportunity. Freeze eligibility, scope, version and exclusions; retain rejected and withdrawn candidates with reasons.
Verified execution	[PD] Require dated evidence linked to the actor, behavior and opportunity. An automatically cleared condition or task-status change does not establish the prescribed act.
Execution quality	[PD] Define the rubric and sampling window. A presence photograph supports a sampled state; it does not demonstrate sustained performance across shifts or sites.
Customer outcome	[PD] Define unique eligible events, overlap, reporting lag and maturity. A changed adoption proxy cannot silently replace an effectiveness endpoint.
Verification workload	[PD] Measure review, exception and rework time against available verifier hours. Counts of templates alone do not establish workload.

[PD] Code unverified, missing, not applicable, rejected and verified failure separately. Publish evidence coverage with the score; never infer zero execution from an empty source-read log. Assess missingness by site, action type and operating pressure. Use best/worst-case bounds and a follow-up sample where nonresponse may depend on performance.

[PD] For taxonomy reliability, stratify an event sample across node types, actions and periods. Two trained reviewers independently code masked records using a frozen handbook before adjudication. Size the sample for a prespecified agreement-interval width. Report raw agreement, prevalence, confusion matrices and Cohen’s κ with uncertainty; use weighted κ only for an ordinal rubric. Repeat calibration after rule changes.

[AI] These protocols are proposed. No observed reliability or denominator-stability estimate is available.

Source S1: proposed scoring template and incomplete action-record observations. Measurement protocol is proposed design.

TECHNICAL FOUNDATION • PERSONAL CONTRIBUTION

Credit the design contribution without converting advice into program ownership

[HR] The register distinguishes advisor-associated design material from proposal authorship, program ownership and field implementation.

Contribution claim	Process-tracing assessment
Design-session association	[HR] An assistant-maintained session-associated record contains intake, capacity and escalation advice. [AI] Association is attested; human origination, delivery and adoption are not uniquely established. No smoking-gun authorship evidence.
Analytical calibration	[HR] The register attributes related verification and timing analyses to the advisor. [AI] Provisional registry-level hoop support; verify originals and scope before stronger attribution.
Tracking integration	[HR] Attested advisor-associated tracking artifacts reference the tiered framework. [AI] This does not establish authorship of that framework or field delivery.
Audit-data sharing	[HR] A linked regional summary assigns the advisor audit-data sharing. [AI] Specific assignment gives stronger registry-level hoop support for this narrow role, not proof of completion.
Adoption and outcomes	[AI] Unestablished. Similar tiers do not distinguish adoption of advice from independent development. No smoking-gun adoption trace or doubly-decisive outcome evidence.

[AI] A hoop test checks necessary evidence; passing leaves alternatives. A smoking gun specifically supports one explanation. A doubly-decisive test also excludes relevant rivals.

[PD] Seek original advice, a decision response and downstream version change. Separate assignment from completion and design from delivery. Another role’s proposal metadata contradicts sole-authorship language.

[AI] The supported claim is an attested advisory association and specific audit-data assignment, with primary verification outstanding.

Source S1: ownership, evidence and chronology records. Names, dates and original metadata remain in the private annex.

TECHNICAL FOUNDATION · SELECTION AND MODERATORS

The gate must not earn a better score by excluding the hardest work

[AI] Action admission, audit selection and cohort graduation can each change the population being judged.

[AI] A strict evidence gate may admit easily observed actions while excluding important but poorly instrumented ones. A completion rate among admitted actions can therefore rise without better execution across the original opportunity set. Preserve the candidate queue, rejection reasons and eligible opportunities; report verified coverage as well as conditional compliance.

Selection or moderator	Design implication
Extreme starting cohort	[PD] Freeze selection before follow-up. Estimate expected reversion using untreated repeat cohorts selected by the same rule; absent those data, present a parameterized sensitivity, not an estimated correction.
Node and business model	[PD] Test whether the action reaches the relevant constraint. Central templates may require local parameters; preserve the common behavioral definition and distinguish permitted adaptation from a new treatment.
Verifier and site capacity	[PD] Assess shared staffing, travel, exception burden and competing duties. A site-level action cap cannot certify regional review capacity without an instance-to-evidence mapping.
Audit and graduation	[PD] Sample difficult, missing and apparently successful records. Retain graduated units in the evaluation cohort; removing them changes the estimand and can conceal relapse.

[PD] Compare focus and non-focus units only with aligned eligibility, calendar periods and maturity. Prespecify baseline moderation and treatment-by-stratum interactions. Differences in subgroup significance are not evidence of different effects.

[AI] The supplied summaries do not identify these selection effects or subgroup responses. Their role here is to define failure modes the prospective pilot must test.

Source S1: focus-cohort design, capacity recommendations and regional implementation summaries. Selection analysis is proposed.

TECHNICAL FOUNDATION • PROSPECTIVE EVALUATION

Test the assurance protocol against independently observed execution, not its own dashboard

[PD] Randomize rollout order at the level that contains shared staff and spillovers.

[PD] A prior paired shadow test estimates gate error on frozen records; it does not establish policy ITT.

[PD] Preregister a candidate-action set, eligible opportunities, proof rubric and authority boundaries. Randomize clusters within baseline strata to phased protocol access. Use sites only when staff and practices are separable; otherwise cluster at the shared operating unit. Retain assignment regardless of admission, compliance or graduation.

Protocol element	Decision requirement
Primary assurance endpoint	[PD] Independently adjudicated unsupported-completion incidence across the fixed eligible opportunity set. Blind the validation audit to operational completion status where feasible.
Coverage and burden	[PD] Track verified execution coverage, rejected/unobserved opportunities, verifier hours, repeat barriers and displaced operating work. A lower false-closure rate alone cannot justify shrinking useful execution.
Customer endpoint	[PD] Measure mature customer harm after prespecified action-response and reporting lags. Treat it as a separate endpoint; evidence capture is not a customer benefit.
Identification	[PD] Account for cluster assignment, calendar effects and repeated observations. Check anticipation, missingness, contamination and changing treatment versions.

[PD] If randomization is infeasible, justify a staggered comparison using not-yet-treated or never-treated clusters and group-time effects. Inspect pre-trends, overlap and composition; a nonsignificant pre-trend test does not establish parallel trends.

[PD] Design-specific power must reflect cluster counts, rollout timing, serial correlation, outcome prevalence, unequal sizes and audit error. Compare MDE with break-even in the same endpoint and horizon. If the feasible design cannot resolve the decision, change the rollout or measurement before launch. The numerical planning analysis is separate from historical evidence.

Proposed evaluation. Methods: randomized phased rollout or justified group-time DiD; refer to verified methodological sources in the references.

TECHNICAL FOUNDATION · LITERATURE AND PROPOSITIONS

An Evidence-Carrying Action Contract makes assurance promises testable

[PD] The proposed framework connects admission, proof, verification capacity, timing and accountable escalation.

[AI] Campbell and Goodhart motivate testing whether an auditable score displaces the underlying objective. Ittner and Larcker motivate validating assumed links from nonfinancial measures to value. Tucker and Edmondson distinguish learning from repeated operational workaround. Shewhart and Deming caution against reacting to variation without a stable interpretation. These works motivate the design; this case does not empirically confirm their causal mechanisms.

[AI] The case extends the practical design question: a proof source requires an action unit, verification owner and available capacity. This is a testable synthesis, not established theoretical novelty.

[PD] P1 · Evidence. Relative to ordinary tracking, actor-linked proof reduces independently adjudicated unsupported closure over the same eligible opportunities. A precise failure to achieve the preregistered decision-relevant reduction disconfirms the proposition.

[PD] P2 · Capacity. Where review demand exceeds staffed capacity, a capacity-matched action portfolio produces greater verified opportunity coverage than an equal-intensity site cap. Test workload and coverage jointly; equivalent coverage without additional burden disconfirms the predicted advantage.

[PD] P3 · Timing. Separating execution and outcome clocks reduces premature action retirement without worsening mature customer harm. A controlled comparison finding no decision-relevant reduction in premature retirement challenges the timing rule.

[PD] P4 · Learning. Routing recurring systemic barriers to a role with corrective authority reduces repeated unresolved barriers compared with returning them to the same site. No meaningful reduction under an implemented escalation policy disconfirms the claim.

[AI] All propositions require prespecified margins, verified implementation and suitable counterfactuals; a noisy null result is not automatic falsification.

Proposed Luna Sol synthesis. Literature supports interpretation and study design, not historical program impact.

TECHNICAL FOUNDATION • LIMITATIONS

Design evidence cannot establish end-to-end adoption or historical impact

[AI] The following limits determine which conclusions this case can support.

Lineage and attribution

[HR] The registry combines metadata, meeting-summary transcriptions, analyses and an assistant-maintained session record. [AI] Originals have not been authenticated within this case. Repeated summaries may share one lineage. Session association cannot uniquely establish human origination, delivery or adoption of advice.

Implementation and missingness

[HR] Framework alignment and regional review activity are recorded, while the complete scoring feed and dashboard were not recovered. A source-read action log was incomplete. [AI] These observations establish an assurance gap; they do not establish zero execution, no local adoption or the cause of missing records.

Capacity, timing and causality

[AI] Regional and site caps may count different action units. Verification hours, reusable evidence and sampling coverage are unspecified; overload is conditional, not measured. Predictive timing from a related analysis does not establish causal effect or transport unchanged to every action. No stable assignment, delivery and mature-outcome panel identifies historical impact.

Economics and alternatives

[HR] No verified benefit or incremental cost is established. [AI] Proposed economics and information value are decision scenarios. Narrowing local actions, improving capability or increasing resolution capacity remain alternatives; this record does not prove centralization uniquely superior or capacity expansion futile.

Transfer and public replication

[AI] Node heterogeneity, shared staff and changing forum scope limit transfer. Public indexing protects internal levels but limits direct aggregation. The private annex preserves source locators, mappings and calculation inputs; arithmetic replication cannot authenticate inaccessible originals. The case supports a bounded design decision, not customer-benefit or savings credit.

Source S1 and this case's evidence audit. Limits apply to every executive and technical claim.

TECHNICAL FOUNDATION · REPRODUCIBILITY AND DEFINITIONS

Freeze the action unit before calculating a score, capacity requirement or benefit

[PD] A reproducible assurance system needs a definition ledger as well as a numerical ledger.

Ledger	Minimum reproduction contract
Source and lineage	[PD] Record hash, version, retrieval status, exact locator and whether the statement was read directly, transcribed or inferred. Keep identity and original-title mappings private.
Action definition	[PD] Link template, site-action instance, eligible opportunity, package version, assigned role, authority and planned execution window. Preserve rejection and withdrawal records.
Evidence and states	[PD] Store actor/event timestamps, proof reference, verification rule, reviewer result, missingness reason, conflict resolution and state history. Do not overwrite a failure with the latest closure status.
Workload and clocks	[PD] Log proof reuse, sampling, review and rework time, verifier availability, action-response lag and customer-outcome closure lag. Separate measurement time from event time.
Claims and calculations	[PD] Map every displayed value to raw input, transformation, assumption, version alternative, formula, rounding and permitted interpretation. Rerun unit and denominator sensitivities.

[PD] Publish protected relative or indexed internal values with a consistent basis. Do not expose both complementary rate movements or a raw-value anchor that reverses indexing. Keep raw sources, confidential parameters and executable calculations in the separate private annex.

[PD] Tag historical record [HR], recalculated [RC], analytical inference [AI] and proposed design [PD]. If a record is unavailable, preserve that status instead of treating it as a failed action. Reconcile conflicts before publication and retain alternatives that change the decision.

[AI] Reproduction demonstrates how the case reached its conclusions. It does not independently verify historical implementation or authenticate source transcriptions.

Proposed reproduction contract. Public source IDs link to a private, versioned source and claim register.

TECHNICAL FOUNDATION • VARIATION AND SELECTION SENSITIVITY

Missing observations cannot support control limits or a measured mean-reversion correction

[AI] This redesign has no observed execution-score time series. Its dates, stages and proposed thresholds are not repeated performance observations.

[PD] For a future commensurate ordered series, compute adjacent $MR_t = |x_t - x_{t-1}|$. Tentative individual limits are $\bar{x} \pm 3 \times \text{mean}(MR)/1.128$; the moving-range upper limit is $3.267 \times \text{mean}(MR)$. Show sample size and phase changes. Fewer than two observations yield no moving ranges; very short series, zero variation or autocorrelation invalidate confident signal classification.

[PD] For binary events, retain event counts and eligible denominators in rational time subgroups; do not treat individual binary records as continuous XmR observations. For large-denominator rates with extra-binomial variation, assess Laney p' : $z_i = (p_i - \bar{p})/\sqrt{[\bar{p}(1-\bar{p})/n_i]}$, $\sigma z = \text{mean}(MR(z))/1.128$; limits are $\bar{p} \pm 3\sigma z\sqrt{[\bar{p}(1-\bar{p})/n_i]}$, bounded to the rate support. Investigate clustering, mix and denominator change before interpreting dispersion.

[AI] No historical signal is classified here. Future within-limit movement is not established improvement; an out-of-limit movement still does not identify a cause.

Synthetic reliability r	Expected reversion from selected excess d = 20
0.30	14 normalized points
0.60	8 normalized points
0.90	2 normalized points

[PD] This illustration assumes a stationary classical Gaussian repeat-measure model: expected reversion is $(1-r)d$. The selected excess and reliability values are invented planning inputs, not internal observations or estimated risk. Focus and non-focus repeated untreated data, aligned definitions and stable selection are needed to calibrate actual reliability and counterfactual recovery.

[AI] The illustration shows sensitivity to reliability; it does not subtract an assumed correction from an unavailable project effect.

Proposed methods and synthetic selection illustration. No historical execution series or estimated treatment effect is supplied.

DECISION MODEL / TWO ERROR RISKS

A gate must prevent unsupported claims without suppressing valid work

[PD] Every quantity on these four pages is synthetic. The fixed denominator is all preregistered eligible action opportunities.

[PD] This completion-evidence decision is distinct from action-intake approval. False authorization accepts unsupported execution evidence; false hold rejects independently supported execution. Measure both across the same upstream population; a narrower admission denominator can hide rejected work.

Scenario [PD]	FA reduction	FH increase	Gate cost	Net value
Low	2.00 points	4.00 points	1.20	0.00
Base	4.00 points	3.00 points	1.00	2.40
High	6.00 points	1.00 points	0.80	5.00

[PD] Normalize harm from one false authorization to 100 value units and one false hold to 20. Net value per eligible opportunity equals $100 \times \text{FA reduction} - 20 \times \text{FH increase} - \text{incremental gate cost}$. The table reports absolute risk differences in percentage points; formulas use proportions.

[RC] In the base scenario, a 4-point FA reduction, a 3-point FH increase and cost 1 produce net value 2.40. Break-even requires a 1.60-point FA reduction. That threshold changes if false holds, delay or displaced operating work cost more.

DECISION RULE

[PD] Use the paired shadow audit to estimate both error risks and effort. Release only after a prospective policy test clears the economic threshold and pre-specified false-hold, capture and workload guardrails. No customer savings are claimed.

Synthetic model D1; no source metric or realized value. Complete joint error distribution and cost reconciliation in reproducibility code.

DECISION MODEL / COST SENSITIVITY

A favorable gate result can disappear when avoidability falls

[PD] This grid prices the same false-authorization risk reduction per fixed eligible opportunity.

[PD] Net decision value = $100 \times a \times (1-o) \times \text{FA reduction} - 20 \times \text{FH increase} - C$. Here a is avoidable error-loss share; o is economic value credited elsewhere; C is incremental gate cost. Hold FH increase at 0.03, C at 1 and o at zero.

FA reduction [PD]	a = 50%	a = 75%	a = 100%
1.60 points	-0.80	-0.40	0.00
3.00 points	-0.10	0.65	1.40
4.00 points	0.40	1.40	2.40

[RC] A 3-point reduction yields -0.10 units at 50% avoidability and +1.40 at full avoidability. With full avoidability, adding 25% overlap raises break-even from 1.60 to 2.13 points. Thus evidence of fewer unsupported claims alone cannot certify positive net value.

[PD] The 100-unit basis contains potential decision-error loss only. Use $a = 1$ if already avoidable and $o = 0$ if already exclusive. Record incremental verification, false-hold, delay and displaced-work costs once. Do not add EVSI or customer benefits to this ledger.

[PD] Customer economics requires a separate mature-outcome ITT: $\text{net} = V \times \text{deltaY} \times k \times aC \times (1-oC) - K$. V is eligible horizon volume; deltaY an absolute mature-defect risk reduction; k loss per defect; aC avoidability; oC overlap; K incremental total cost. For relative reduction r , $\text{deltaY} = \text{untreated mature risk} \times r$. These actual inputs are unknown. Define and power that customer trial before claiming impact.

Synthetic model D1S; all grid cells and overlap thresholds are reproduced in code. Customer-value contract is parameterized, unsized and unpowered.

DECISION MODEL / SHADOW PRECISION

Nominal audit volume can overstate decision precision

[PD] A shadow comparison holds underlying cases and independent reference truth fixed. It tests decision quality, not customer impact.

Independent pairs [PD]	Exact MDE	Power at 1.60 points	Precision
1200	2.35 points	46.17%	Insufficient
3200	1.43 points	88.39%	Conditional

[PD] Assume 8% paired discordance. Conditional on discordant count M , beneficial reversals follow $\text{Binomial}(M,q)$, $q=(0.08+\delta)/0.16$. Integrate the conservative exact two-sided test at α 0.05 over $M \sim \text{Binomial}(n,0.08)$; solve for 80% power. This tests zero difference, not whether a lower confidence bound clears break-even.

[RC] With 3,200 nominal pairs,20 pairs per cluster and paired-difference ICC 0.02, a variance-screening design effect of 1.38 gives an approximate MDE of 1.65 points—above 1.60. This heuristic is not exact cluster power. The independent-pair result therefore cannot certify a clustered field design.

[PD] Next randomize the gate policy at shared decision-authority level. Estimate ITT unsupported-completion incidence on fixed eligible opportunities with a preregistered closure clock and blinded ascertainment. Retain unresolved opportunities; coverage and delay guardrails prevent deferral from appearing successful. Use pilot covariance and cost to simulate field power. Mature customer impact remains unsized and unpowered.

Synthetic model D2; exact paired-binomial enumeration within numerical tolerance $1e-10$. Cluster adjustment is approximate and not a field sample-size commitment.

DECISION MODEL / VALUE OF THE NEXT AUDIT

An informative study can still be a poor investment

[PD] This separate Bayesian model values a future audit that can change whether the gate package is released.

[PD] Release yields+20 aggregate decision units if the package is viable and−40 otherwise; hold yields 0. Release only above posterior viability 2/3. The audit costs 2 units. These payoffs have a different normalization from the per-opportunity error ledger.

Scenario [PD]	Prior viable	Sensitivity / specificity	EVSI	Net study
Low	0.80	0.65 / 0.65	0.00	-2.00
Base	0.60	0.80 / 0.80	6.40	4.40
High	0.60	0.95 / 0.95	10.60	8.60

[RC] In the low scenario, the preferred action remains release after either audit result: EVSI is 0 and the study destroys 2 units after cost. In the base scenario, hold becomes release only after a pass; expected information value is 6.40 and net study value 4.40.

[RC] EVSI equals expected best posterior payoff minus best prior payoff. Every scenario satisfies $0 \leq \text{EVSI} \leq \text{EVPI}$; perfect-information bounds are 8,12 and 12 units. Net study value subtracts study cost separately.

[AI] These calculations demonstrate the evidence needed to value a decision; they do not price a historical correction. Audit accuracy is assumed, not inferred from the power table. Add delay cost where material, and never add these aggregate decision units to customer benefits or the preceding error-cost ledger.

Synthetic model D3; exact finite-state Bayesian enumeration, using the EVSI definition in Strong et al. (2015). Scenarios are not confidence intervals.

The case makes its evidence boundary explicit

[AI] Historical statements are register-attested. Proposed calculations are transparent decision scenarios.

Label	Meaning
[HR] Historical record	The supplied project register reports a statement, attribution or artifact. It does not mean an unavailable original was independently authenticated.
[RC] Recalculated	A deterministic transformation of supplied values or a clearly identified synthetic scenario. The input class is stated with the result.
[AI] Analytical inference	An interpretation of the record or a methodological implication. It is not an identified historical effect.
[PD] Proposed design	A future intervention, study protocol, decision rule or synthetic assumption. It is not claimed as deployed or observed.

[HR] S1 is the supplied master project register. It contains a project narrative and linked metrics, decisions, implementation records, stakeholder roles, source references and chronology. E, M and D codes identify evidence, metric and decision rows within the selected project; full identifiers, worksheet rows, cells and original names are private. The source workbook was not modified.

[AI] Executive titles inherit the evidence scope of their exhibit. The cover is an analytical decision statement; its finding is an inference from conflicting action units, not measured workload. Public scenario parameters, sample counts and methodological constants are not internal operating metrics.

[PD] The private reproduction package includes the selected extract and hash, claim ledger, source audit, raw conditional capacity calculations, executable synthetic model, model outputs and rendering source. The public calculations file contains only synthetic assumptions and derived outputs. Raw operating levels and identity mappings remain outside the public case.

[AI] Reproduction can confirm arithmetic and provenance mapping. It cannot authenticate absent original files, prove that recorded proposals were adopted or establish historical effects from an unobserved evaluation.

Evidence classification applies throughout. The private package is for the owner’s evidence review and must remain separate from the portfolio PDF.

REFERENCES / IMPLEMENTATION, FLOW AND MEASUREMENT

Execution assurance combines behavioral and measurement disciplines

Primary methodological sources. They support the analytical design and interpretation; none establishes this project's results.

R1. Carroll, C., et al. (2007). A conceptual framework for implementation fidelity. *Implementation Science*, 2, 40. Content, coverage, frequency and duration are distinct fidelity dimensions.

R2. Little, J. D. C. (1961). A Proof for the Queuing Formula: $L = \lambda W$. *Operations Research*, 9(3), 383–387. A flow identity under stated conditions; it is not a staffing or deadline guarantee.

R3. Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3), 581–592. Missingness assumptions govern when incomplete records can support inference.

R4. Srigley, J. A., et al. (2014). Quantification of the Hawthorne effect in hand hygiene compliance monitoring using an electronic monitoring system: a retrospective cohort study. *BMJ Quality & Safety*, 23, 974–980. Motivates checking observer reactivity; no effect magnitude is transported.

R5. Hernán, M. A., & Robins, J. M. Causal Inference: What If. Author-hosted text. Treatment definition, causal diagrams, potential outcomes and identification.

R6. NIST/SEMATECH. e-Handbook of Statistical Methods: Individuals Control Charts. Moving-range estimation and individuals limits.

R7. Laney, D. B. (2002). Improved Control Charts for Attributes. *Quality Engineering*, 14(4), 531–537. Attribute-chart adjustment for extra-binomial variation.

References were checked against publisher, author or institutional records. Linked entries distinguish public methods from private project evidence.

REFERENCES / EVALUATION AND INFORMATION VALUE

The proposed tests must resolve a specified operating decision

Primary methodological sources. They support the analytical design and interpretation; none establishes this project's results.

R8. Callaway, B., & Sant'Anna, P. H. C. (2021). Difference-in-Differences with Multiple Time Periods. *Journal of Econometrics*. Author manuscript on group-time effects and valid comparison groups.

R9. Roth, J. (2022). Pretest with Caution: Event-Study Estimates after Testing for Parallel Trends. *American Economic Review: Insights*. A nonsignificant pre-trend test is not proof of parallel trends.

R10. Hussey, M. A., & Hughes, J. P. (2007). Design and analysis of stepped wedge cluster randomized trials. *Contemporary Clinical Trials*, 28, 182–191. Schedule-specific cluster design and analysis.

R11. Hemming, K., et al. (2015). The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting. *BMJ*, 350, h391. Rollout, period effects and implementation constraints.

R12. 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*. Information value is the gain from a better decision under uncertainty.

R13. Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*. Chance-corrected independent classification agreement.

R14. Collier, D. (2011). Understanding Process Tracing. *PS: Political Science & Politics*. Hoop, smoking-gun and doubly-decisive evidence tests.

R15. 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. Extreme-cohort selection and repeated measurement.

References were checked against publisher, author or institutional records. Linked entries distinguish public methods from private project evidence.

REFERENCES / INCENTIVES AND ORGANIZATIONAL LEARNING

A control system should reveal problems rather than reward their disappearance

Primary methodological sources. They support the analytical design and interpretation; none establishes this project's results.

R16. Campbell, D. T. (1976). Assessing the Impact of Planned Social Change. Original occasional-paper record, ERIC. Measures under evaluative pressure can distort the process they represent.

R17. Goodhart, C. A. E. (1984). Problems of Monetary Management: The UK Experience. In *Monetary Theory and Practice*, pp. 91–121. Publisher chapter record and abstract; a conceptual analogue for unstable indicator relationships under control.

R18. Ittner, C. D., Larcker, D. F., & Randall, T. (2003). Performance implications of strategic performance measurement in financial services firms. *Accounting, Organizations and Society*. Validating nonfinancial performance relationships.

R19. Tucker, A. L., & Edmondson, A. C. (2003). Why Hospitals Don't Learn from Failures: Organizational and Psychological Dynamics That Inhibit System Change. *California Management Review*; author working paper. Workarounds and system-level learning.

R20. The W. Edwards Deming Institute. The Funnel Experiment. Institutional explanation of tampering and common-cause variation.

R21. OECD / European Commission JRC (2008). *Handbook on Constructing Composite Indicators: Methodology and User Guide*. Weighting, normalization, comparability and sensitivity.

R22. Wheeler, D. J. (2012). *Exact Answers to the Wrong Questions*. SPC Press. Process behavior and the interpretation of statistical variation.

References were checked against publisher, author or institutional records. Linked entries distinguish public methods from private project evidence.

FINAL REFEREE REPORT / AUTHOR-LED, AGENT-ASSISTED

The case supports a bounded operating-design decision

[AI] Internal editorial review assessed executive usefulness, evidentiary validity and numerical reproducibility. This is not external journal peer review.

Major issue	Revision and disposition
Cap arithmetic	Retained alternative unit mappings and staffing interpretations privately. Public wording is qualitative; paired multipliers would reveal an internal footprint. No actual overload is inferred.
Missingness	Removed a zero-execution baseline and the inference of no adoption. Preserved the observed empty extract alongside regional activity.
Personal attribution	Separated assistant-maintained advisory-session memory, a narrow recorded assignment and others’ authorship. Similarity of ideas is not an adoption trace.
Causal and metric scope	Aligned executive and technical diagrams. Kept related coefficients, different metric definitions, clocks and estimands out of the redesign’s results.
Decision economics	Separated synthetic gate-decision value from customer impact. Detectability and break-even refer to the same error contrast; the live rollout needs its own powered protocol.

[AI] Minor revisions harmonized the action-instance vocabulary, evidence labels and single-snapshot wording. The cover now identifies the project type, states the decision and bounds the contributor role. Sources are linked; raw evidence and source-version alternatives remain private.

[AI] Referee disposition: acceptable as a register-based operating-design and evaluation case within its stated boundaries. Historical adoption, customer impact, realized savings, source authentication and final field-test power remain unresolved. These are not converted into negative findings.

[AI] Partner critique: a COO can act on page one—define the verification unit, instrument workload and fund a bounded gate-validation pilot. The new finding changes the approval question: action-count limits cannot certify verification capacity until their units are reconciled.

[PD] Reopen the scale decision when the action ledger, independent adjudication, observed verification effort and policy-test feasibility are established. Reopen customer-value credit only after mature outcomes and finance-approved costs support it.

Three parallel review workstreams informed revision. Final delivery includes executable calculations, confidentiality screening, layout checks and rendered-page inspection.