

OPERATING CASE STUDY / OPERATIONAL EXCELLENCE

Recovery Forecasting & Benefit Attribution

A booked commitment is not verified performance

An anonymized project case study of a service-network reforecast: reconciling intervention bookings, isolating model revisions and defining the evidence required to turn an operating commitment into defensible performance credit.

PROJECT CONTEXT / OPERATIONS, ANALYTICS & IMPLEMENTATION

BLUF / BOTTOM LINE UP FRONT

Keep the scenarios visible.
Verify the recovery credit.

RECALCULATED FINDING

The positive forecast cushion equals only 2.71% of the whole-save upward revision. Other simultaneous edits prevent assigning that revision to one intervention. A controlled replay must precede computational credit; outcome evidence must precede performance credit.

CONTRIBUTION IN THE SOURCE RECORD

Reforecast modeling, analytical challenge and dual-scenario communication. Registry-attested contribution; provisional hoop-level support.

EVIDENCE STATUS

Planning and evidence-reconciliation case. Original model authentication, causal customer impact and realized savings remain unestablished.

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

EXECUTIVE ABSTRACT

Verify the recovery credit before relying on the forecast

A service-network reforecast paired a booked recovery commitment with a downside scenario. Its decision value depends on distinguishing model edits, delivery evidence and validated outcomes.

CONTEXT | A commitment changed the represented forecast

[HR] The register describes a named intervention booking, a positive committed case and a negative stress case. Several other cells changed in the same saved workbook. Later versions retain the booking while revising the overall position.

OBJECTIVE | Establish what the booking can legitimately claim

[PD] Separate a computational contribution to a forecast from authorization, delivered work, an identified customer effect and economic value. Each requires its own evidence and accountable owner.

APPROACH | Reconcile the increment, replay the edits, type the evidence

[RC] Reconstruct the whole-save revision and compare the narrative increment with registered cells. [AI] Audit scenario and actuals interpretations. [PD] Define a Forecast Credit Protocol, with controlled replay, evidence states and prospective evaluation.

FINDINGS | The favorable position has little room for revision

[RC] The positive cushion equals 2.71% of the whole-save upward revision. The later-month booked increment is 196.85 on a scale where the narrative increment is 100. [AI] Neither comparison isolates the intervention; later forecasts cannot validate an earlier forecast against reality.

DECISION VALUE | Authorize verification before performance credit

[PD] The COO should appoint a forecast-credit owner and commission a controlled replay. Require the model steward to reconcile the increment and dependencies, the operating owner to supply delivery evidence, and Finance to approve entitlement and overlap rules.

IMPLICATIONS | Credit the analysis within its evidence boundary

[AI] Modeling and dual-scenario communication have provisional registry-level hoop support. Original workbooks and sent communications were not authenticated. The case establishes neither isolated booking contribution nor historical customer impact or realized savings.

S1: selected project register, M01–M13 and linked evidence. HR historical record; RC recalculated; AI analytical inference; PD proposed design. Public historical values are indexed or relative; original artifacts were not recovered.

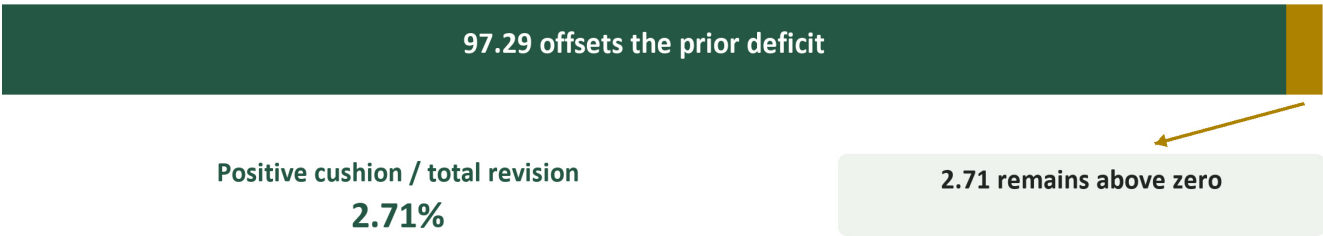
EXHIBIT 1 / FORECAST HEADROOM

The positive cushion is only 2.71% of the whole-save revision

[RC] A small reversal of the recorded upward revision would eliminate the favorable forecast position. This is a sensitivity of the model, not an estimate of intervention impact.

ALMOST ALL OF THE REVISION OFFSETS THE EARLIER DEFICIT

Whole-save upward change = 100 • endpoint arithmetic • no isolated program effect



An adverse revision equal to the cushion removes the positive position.
This is not a failure probability or a demonstrated customer outcome.

[HR] The registered save changes a negative modeled position to a positive one. [RC] Setting the full upward change to 100 leaves a positive cushion of 2.71; the remaining 97.29 offsets the earlier deficit. An adverse revision equal to the cushion reaches zero, holding other values fixed.

Interpretation	Decision boundary
[RC] Revision sensitivity	The denominator is the whole-save model change. The figure is neither a probability of failure nor a share of verified program benefit.
[AI] Credit attribution	The save includes other edits. The named intervention cannot receive the entire revision merely because its booking appears in that file.

[AI] A positive endpoint can be arithmetically correct and operationally unsupported. An unchanged booking in later versions establishes persistence in the represented model, not durable performance. The cushion therefore increases the importance of checking dependencies and uncertainty before using the forecast to reduce management attention.

COO ACTION

[PD] Keep the committed and downside cases visible. Ask the forecast owner to identify which assumptions can erase the cushion, the evidence behind them and the threshold that reopens approval. Do not convert the cushion into a confidence statement.

[HR/RC] S1:M01/M03/E03/E04. Whole-save upward revision indexed to 100; zero position retained. Registered pre/post endpoints determine the cushion. The save is not a controlled intervention comparison.

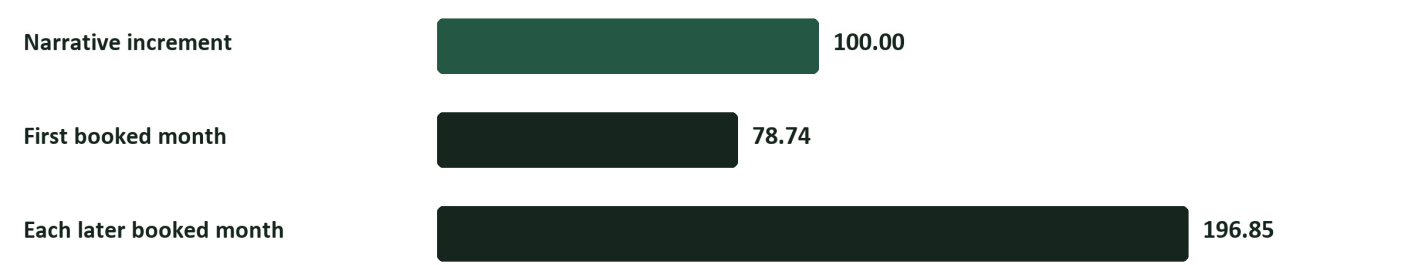
EXHIBIT 2 / BOOKING IDENTITY

The narrated increment does not reconcile to the recorded booking

[RC] The later-month cell increment is 196.85 when the narrated increment is indexed to 100. The first future month carries a different increment.

THE NARRATIVE AND THE BOOKING NEED A METRIC CROSSWALK

Narrative later-month increment = 100 • registered changes • compatible scopes unverified



A rate-to-unit bridge needs population, exposure, month alignment and overlap rules.
The mismatch is unresolved; different scopes could explain it.

[HR] The narrative describes a common later-month recovery increment. The registered program row and month-aligned summary carry a larger change in those months and a smaller first-month change. [AI] A shifted program-row layout makes period alignment a validation task, not a formatting detail.

Reconciliation requirement	Evidence needed
[PD] Common metric contract	Confirm population, baseline, outcome, unit, applicable month and whether the increment is gross, incremental or net of another credit.
[PD] Rate-to-unit dependency	Trace each booked rate through eligible exposure to the credited-unit output. Explain why the companion plan-units row is unchanged.

[AI] The mismatch is not yet proof of overbooking: the quantities may have different scopes or baselines. Nor does an unchanged companion row prove a broken formula. Both require the original dependency graph. Adding monthly rate changes produces rate-months; it cannot establish unique customer events prevented.

VALIDATION HOLD

[PD] Ask the model steward to supply a signed metric crosswalk and dependency replay. Retain both representations until reconciled. Do not select the narrative or the cells solely because one produces a more favorable result.

[HR/RC/AI] S1:M02/M07/E03/E04. Narrative increment =100; first and later booked increments =78.74 and 196.85. These protected comparisons require a common metric contract before substantive interpretation.

EXHIBIT 3 / COMPUTATIONAL ATTRIBUTION

Two saved endpoints cannot isolate one booking from a bundle of edits

[AI] The observed save changes the named booking and other model inputs. Recover the missing replay states before assigning computational credit.

RECOVER THE MISSING REPLAY STATES BEFORE ALLOCATING CREDIT

Proposed controlled replay • B = booking; O = other edits • freeze the calculation engine

	Old booking B0	New booking B1
Old other edits O0	f00: recorded endpoint Original engine unverified	f10: missing replay Change booking only
New other edits O1	f01: missing replay Change other edits only	f11: recorded endpoint Original engine unverified

Booking response: f10 – f00 at O0; f11 – f01 at O1.

Interaction = f11 – f10 – f01 + f00. Allocating it is a convention.

A computational bridge does not identify a customer treatment effect.

[PD] Replay old and new booking states against old and new other-edit bundles under one frozen calculation engine. Compare the booking at each bundle state and retain their interaction. The observed diagonal contrast alone cannot recover these missing marginal contributions.

Replay control	Required evidence
[PD] Fixed engine and data	Preserve formulas, calculation mode, linked data, actuals cut, exposure, hidden ranges, month mapping and rounding. Recalculate rather than read caches alone.
[RC] Correct edit vector	A blanket verbal description of another row overstates its corrected net change by 0.63%. The period-specific registered vector reconciles to the annual change.

[AI] Even a complete replay identifies a model response, not a customer response. Dividing interaction between edit groups is an allocation convention. It cannot certify uniquely correct ownership or causal benefit. If combinations are incoherent or dependencies change, expand the replay and leave the unexplained residual unallocated.

ATtribution Rule

[PD] Release computational credit only when a reproducible edit ledger closes within a declared tolerance. Keep operational and financial approval separate. Missing replay states are unresolved evidence, not zero contributions.

[HR/RC/AI/PD] S1:M02/M07/M08/M12/E03/E04. Corrected net edit indexed to 100; blanket-vector alternative =100.63. Four-state replay is proposed; no original engine or off-diagonal result was recovered.

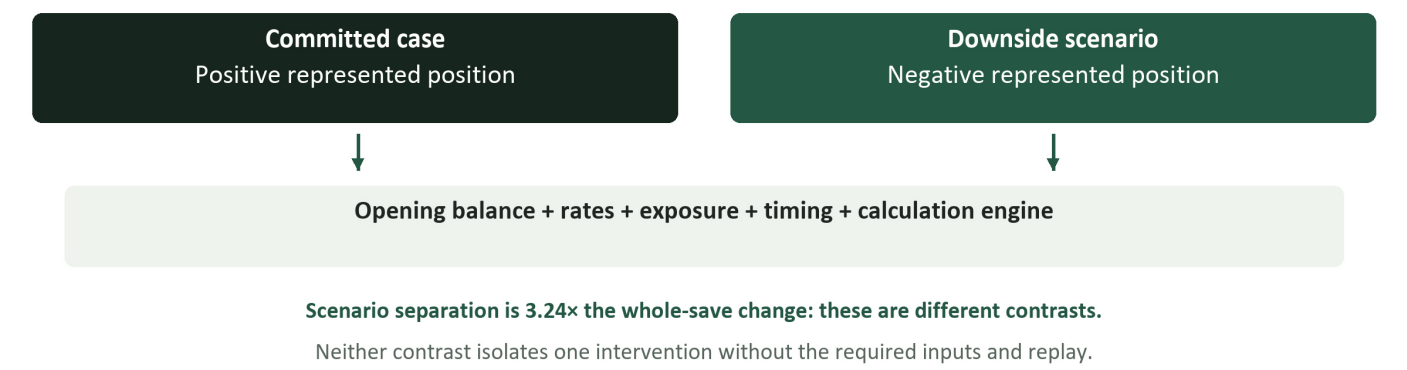
EXHIBIT 4 / SCENARIO DISCIPLINE

A disclosed downside case still needs a reproducible model contract

[AI] Presenting alternatives makes uncertainty visible. It does not validate either endpoint or establish that their spread is the named intervention’s contribution.

A SCENARIO ENDPOINT NEEDS A REPRODUCIBLE CONTRACT

Recorded disclosure • conditional alternatives • no calibrated probability distribution



[HR] The draft pairs a committed case with a negative run-rate scenario. Its supporting schedule has unresolved unit and attainment semantics; the endpoint cannot be reproduced from the recovered register alone. [AI] The spread compares scenarios whose other inputs are not proven identical.

Source interpretation	Required correction
[AI] Higher recent average	A trailing-period average and an overlapping cumulative average do not establish a monthly acceleration pattern or validate the future path.
[AI] Full monthly attainment	Meeting one period’s target need not erase earlier cumulative deficits. That claim cannot accept or reject the stress endpoint without its full ledger.

[PD] Reconstruct the opening position, future rate assumptions, exposure, attribution rules and cumulative sign convention. Declare whether each endpoint is a conditional scenario, a forecast or an authorized commitment. Two scenarios without calibrated probabilities are not a probability distribution or a confidence interval.

DECISION USE

[PD] Use the downside case to identify conditions that require action. Mark its unreproduced inputs and request reconciliation. Preserve disclosure while withholding validated forecast accuracy and isolated intervention-credit claims.

[HR/AI/PD] S1:M04/M05/M06/D02/E01/E02. Scenario arithmetic and substantive validation are separate. Neither the stress schedule nor a historical probability distribution is authenticated.

EXHIBIT 5 / EVIDENCE STATES

A zero in the actuals row does not identify a zero intervention effect

[HR] The represented program-actual row remains all zero before and after booking. [AI] Its evidence state must be resolved before interpreting it as performance.

RECORD THE EVIDENCE STATE BEFORE INTERPRETING ZERO

Proposed recording taxonomy • classify the source cell before granting performance credit

Observed zero Eligible outcome measured	Default Template value retained	Unrecorded Expected evidence absent
Not due Maturity window open	Inapplicable Outside the metric scope	Imputed Rule-derived estimate

The same displayed value can represent materially different evidence.

Zero-filled actuals do not distinguish delivered effect from missing measurement.

[AI] The row supports one narrow statement: no nonzero attributed actual is recorded in that representation. A default, missing value, not-due observation, inapplicable field, imputation and validated observed zero are different states. A zero-filled plan is not a measured monthly outcome panel.

Credit claim	Minimum evidence
[PD] Work delivered	An eligible action record, accountable owner, completion evidence and independent verification of the prespecified package.
[PD] Outcome improved	A fixed eligible population, mature customer outcome and valid treatment comparison. Delivery or an attributed spreadsheet entry alone is insufficient.

[HR] The record promises follow-up and preserves the booked plan. [AI] Neither proves that the follow-up occurred or that a measurement system was implemented. An empty actuals row for one intervention also cannot settle entitlement questions concerning different program families. Keep those claims within their own scope.

EVIDENCE CAPTURE

[PD] Require the operating owner to populate typed evidence states and a linked outcome ledger. Independent reviewers should audit both valid zeros and unresolved records. Preserve unresolved status rather than forcing a favorable or unfavorable conclusion.

[HR/AI/PD] S1:M10/D04/E01/E03/E04/E06/E07. Zero is a registered recording state, not a confidential nonzero level. Original formula/cache status and program-outcome measurement were not authenticated.

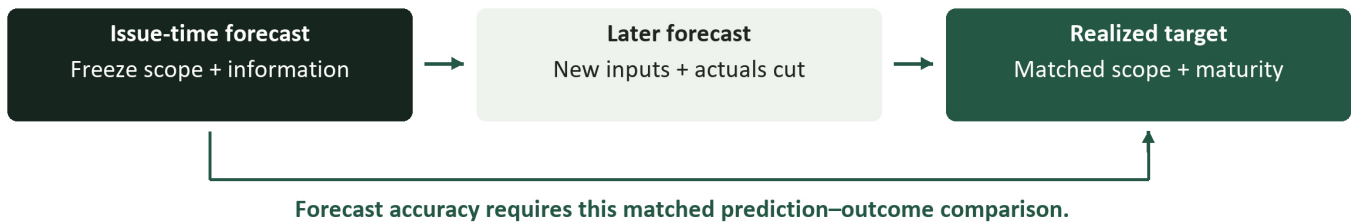
EXHIBIT 6 / FORECAST VALIDATION

A later forecast cannot prove an earlier stress forecast was right

[HR] Subsequent model versions move the overall position while retaining the booking. [AI] Those revisions are not the realized targets needed to score the earlier forecast.

SCORE THE FROZEN PREDICTION AGAINST ITS REALIZED TARGET

Proposed validation sequence • later forecast versions are new predictions, not outcomes



The record contains forecast versions; the final realized target is not recovered.

A directional revision cannot validate the earlier downside scenario.

[AI] A new version can change actualized months, future assumptions, exposure and attribution. Replacing an earlier negative forecast contribution with an actuals convention may improve one component while the remaining forecast worsens. The resulting net revision does not identify an operational effect or explain which assumption failed.

Validation object	Correct comparison
[PD] Forecast accuracy	Freeze the issue-time prediction, target definition, information set and maturity rule; compare with the matched realized target when observed.
[PD] Intervention impact	Compare mature outcomes under assigned action versus a valid control. A forecast revision or directional scenario agreement is not that contrast.

[AI] The same distinction protects contribution claims. Registered modeling and dual-scenario communication have provisional hoop support; the originals are unavailable. Source chronology also does not establish a verified save-then-send sequence. Neither leadership visibility nor repeated file lineage proves sole authorship, approval or field results.

ACCOUNTABLE REVIEW

[PD] Ask the forecast owner to preserve as-issued versions and schedule matched outcome scoring. Ask the operating owner for a separate treatment evaluation. Keep computational, predictive and causal findings in distinct review fields.

[HR/AI/PD] S1:M03/M09/M10/D01–D05/E01–E07. Later files are registered forecast vintages, not realized year-end outcomes. Contribution grade: provisional registry-level hoop; no smoking-gun test passed.

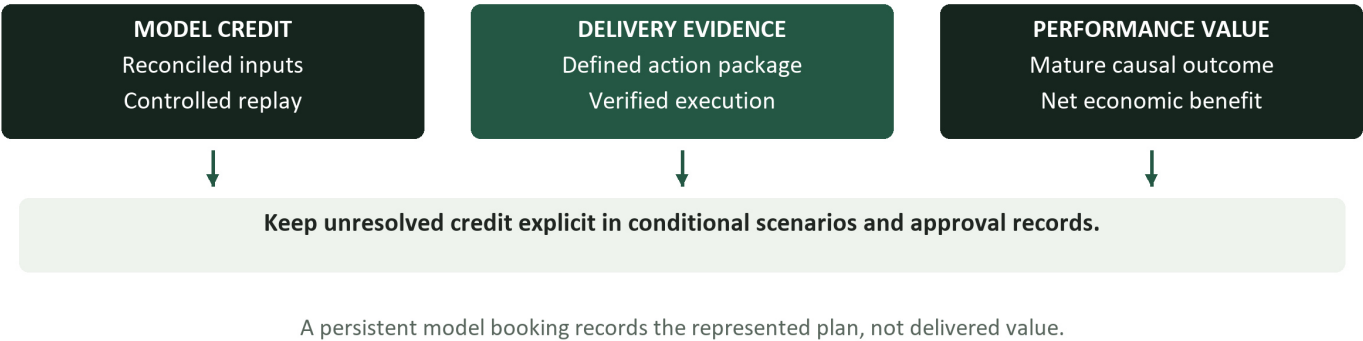
EXHIBIT 7 / FORECAST CREDIT PROTOCOL

Authorize the forecast, delivered work and performance credit separately

[PD] The Forecast Credit Protocol assigns each claim to its own evidence gate. One approved spreadsheet cannot stand in for all three.

THE FORECAST CREDIT PROTOCOL SEPARATES THREE CLAIMS

Proposed governance • one forecast-credit owner coordinates distinct approval gates



Approval gate	Accountable evidence
[PD] Model credit	Model steward reconciles the increment, unit conversion, full edit partition and controlled replay; approver records conditional assumptions.
[PD] Delivery evidence	Operating owner verifies the action package, eligible population, completion and outcome-recording states.
[PD] Performance value	Independent evaluation identifies mature incremental outcomes; Finance resolves entitlement, overlap, realization and incremental cost.

[PD] The COO appoints one forecast-credit owner to reconcile gates and reopen approval after material changes in the model, scope, exposure, delivery or outcome evidence. Unresolved credit remains visible in conditional scenarios, with its status explicit. It does not silently become verified benefit.

COO DECISION

[PD] Authorize the replay and a bounded governance evaluation. Test whether FCP reduces decision burden while preserving approval accuracy, decision coverage and timely resolution. Credit the analytical contribution; withhold unestablished customer and financial results.

[PD] Luna Sol proposed Forecast Credit Protocol, informed by S1:D01–D05 and evidence conflicts. The protocol and prospective test are proposed; historical deployment, customer impact and economic return are not established.

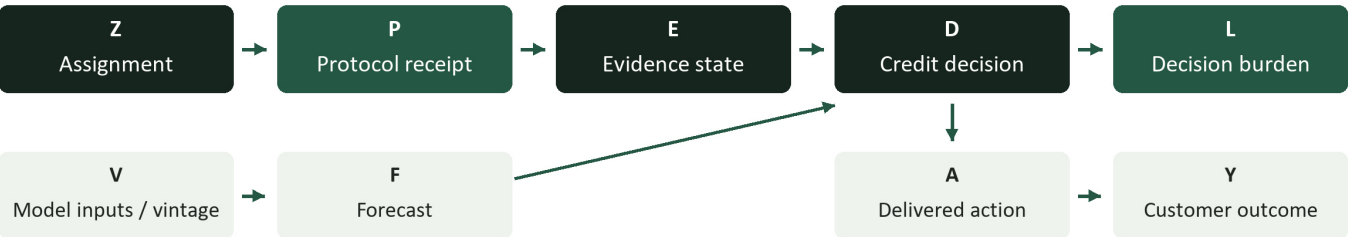
TECHNICAL FOUNDATION • THEORY OF CHANGE

A forecast credit decision is upstream of a customer result

[PD] The Forecast Credit Protocol (FCP) separates model verification, authorization, delivery and outcome validation. None of these stages substitutes for the next.

THE GOVERNANCE TRIAL TESTS DECISION BURDEN

Proposed DAG • historical edge effects unestimated • primary endpoint L at 90 days



Additional paths: P → L; Q measurement quality → E; V → F → D.
C demand, staffing, scope, policy and concurrent change → E, D, L, A and Y.
Z identifies the total effect on L; do not adjust for mediators E or D.
Delivery A → Y needs a separate field design. Residuals do not identify missing causes.

[PD] Z assigns the protocol; P is receipt; E evidence classification; D the credit decision; L decision-resource burden. D may change delivered action A and mature customer outcome Y. Versioned inputs V determine forecast F. Measurement quality Q affects E. Context C includes complexity, workload, analyst experience, budget pressure, demand and concurrent programs.

[AI] Registered artifacts support a proposed F→D pathway and represented model edits. They do not identify D→A or A→Y. The governance trial estimates the total assignment effect Z→L, including changes in review effort and downstream correction. Conditioning on completed reviews, approved credits or adherence would select on protocol consequences.

[PD] P→L captures protocol-induced changes in ordinary review; incremental protocol resources are recorded separately for economics. Historical relationships among C, D and L remain confounded. Residual forecast error can reflect missing inputs, model misspecification or outcome measurement; it does not identify a named cause. Index feedback across review cycles so the diagram remains acyclic.

Source S1: registered forecast edits, scenario communication and unverified actuals. Diagram is a proposed theory of change; historical path effects are unestimated.

TECHNICAL FOUNDATION • ESTIMANDS

Declare which difference the number estimates

[AI] A scenario spread, a revision and a treatment effect answer different questions, even when expressed in the same output units.

Quantity	Estimand and required boundary
Scenario contrast	[PD] Difference between outputs under two fully specified input scenarios at one origin, population and target horizon. It describes the model, not a realized intervention effect.
Whole-save revision	[PD] Output difference between saved states. Booking, other inputs, actuals and formulas may all change. Retain their bundled character until replayed.
Computational booking effect	[PD] Output change when only the booking changes under one engine and fixed other inputs. Its value is conditional on that model contract.
Governance policy ITT	[PD] Control-minus-assignment difference in 90-day ordinary decision-resource burden per upstream-registered decision slot, averaged over eight post weeks.
Field intervention effect	[PD] Mature customer-outcome contrast under delivered action versus a defined alternative in an eligible population. No treatment/outcome comparison identifies it here.

[PD] A slot is a scheduled model-by-forecast-cycle decision, fixed before assignment. Preserve held, withdrawn, unsubmitted and late slots; start follow-up at the original decision date. Eligibility cannot depend on whether the protocol produces a completed review. The trial changes governance, not the field package.

[PD] Benefit uses one pretrial burden reference: $100 \times \text{absolute mean reduction/reference mean}$. Equal slot exposure and common outcome prices underpin the planning model. Team-specific indexing or changed portfolio weights require reweighting and a new power calculation.

Source S1: incompatible scenario, save and attribution comparisons. Definitions are proposed; none reclassifies the register as an identified historical effect.

TECHNICAL FOUNDATION • MEASUREMENT CONTRACT

Make the evidence state explicit before interpreting a cell

[HR] The represented program-actual row contains zeros before and after the booking. Their provenance and semantics are not authenticated.

Measure	Required validation
Booked rate	[PD] Freeze event definition, eligibility, month, denominator and reference entitlement. Distinguish incremental from total rate; resolve shifted columns through an authoritative header crosswalk.
Credited units	[PD] Trace the rate through eligible exposure and dependencies. Recompute outputs using formulas and the actuals cut. An unchanged companion row is a diagnostic, not proof of an unused dependency.
Forecast position	[PD] Identify cumulative versus remaining-horizon logic, sign and actualization rules. A label suggesting year-to-date does not establish its horizon.
Program actual	[PD] Distinguish observed zero, default, unrecorded, not due, inapplicable and imputed. Unknown evidence stays unknown; a default is not an effect estimate.
Decision burden	[PD] Log ordinary review, rework, corrections and nonduplicated delay by fixed prices. Separate incremental protocol cost and preserve unresolved slots in the denominator.

[PD] Proposed taxonomy audit: independently double-code 240 records stratified by version, metric and evidence state. Blind coders to proposed credit; retain prevalence weights when disputed records are oversampled. Report the confusion matrix, raw agreement and Cohen’s kappa with team-bootstrap intervals. Pilot dependence and prevalence before confirmatory sizing; kappa does not establish truth.

[PD] Test invariance across business lines, node types and model versions before pooling. Keep original and harmonized snapshots. Unresolved definitions trigger a release hold with an owner and deadline, not a fabricated zero.

Source S1: rate/unit mismatch, shifted mapping and all-zero representation. Hermans et al. supports dependency inspection; Sargent distinguishes verification from validity.

Two saved endpoints cannot isolate interacting edits

[PD] Define $fbo=F(B=b,O=o;E)$: B is the booking, O other edits, and E the frozen engine/data contract. Indices 0 and 1 denote old and new states; recover all four combinations.

Replay quantity	Exact expression
Booking, old other edits	[PD] $f10 - f00$
Booking, new other edits	[PD] $f11 - f01$
Interaction	[PD] $I = f11 - f10 - f01 + f00$
Symmetric booking allocation	[PD] $\phi B = [(f10 - f00) + (f11 - f01)] / 2$
Symmetric other-edit allocation	[PD] $\phi O = [(f01 - f00) + (f11 - f10)] / 2$; $\phi B + \phi O = f11 - f00$

[AI] The observed diagonal snapshots do not identify either marginal or I. Even their direct comparability requires an authenticated E. The symmetric allocation shares interaction by convention; it is not a unique causal truth. Missing states remain uncomputed. Expand edit groups when a bundled residual is material, and reject impossible combinations.

[PD] Freeze formulas, cached values, calculation mode/engine, external links, data hashes, actuals, exposure, calendar, hidden ranges, month crosswalk, signs and rounding. Recalculate both endpoints first, then change one coherent edit group at a time and complete the factorial replay. Record discrepancies against a declared tolerance.

[AI] A verified computational response establishes how that model allocates output. Field attribution still requires treatment timing, eligible exposure, mature outcomes and a comparison design. Local replay also does not describe the full uncertainty of untested input combinations.

Source S1: simultaneous edits and unavailable original dependencies. Sargent (2020); Sundararajan–Najmi (2020); Ma–Tourani (2020). Equations are proposed replay identities.

Score the original forecast against its original target

[AI] A later reforecast contains a new information set. Agreement with it is not evidence that an earlier scenario predicted the realized year-end result.

Validation record	Contract before observing the result
Origin and inputs	[PD] Archive the issued forecast, formula/data hashes, actuals available then, action assumptions and publication time. Recover a transmitted artifact to reconcile draft/save chronology.
Outcome target	[PD] Specify event, population, horizon, maturity and aggregation. Distinguish realized cumulative outcome from a remaining-period model position.
Comparator	[PD] Freeze a feasible naïve or established forecast using the same issue-time information. Score both against the same mature target; do not choose a comparator after the outcome.
Prediction object	[PD] A point forecast, probability distribution and stress scenario are different objects. A pair of scenarios does not supply calibrated probabilities.
Scoring	[PD] Use a prespecified loss for point forecasts and a proper score for distributions. Inspect calibration and interval coverage across repeated comparable forecast origins.

[AI] When a month is actualized, replacing its prior modeled contribution can change the aggregate mechanically. Preserve a revision bridge separating new actuals, exposure changes, assumptions and engine changes. The favorable removal of an old negative term is not a measured favorable customer effect.

[AI] One future month meeting its rate goal can coexist with an unresolved earlier cumulative deficit. Test the rate-to-unit accumulation equation before interpreting monthly attainment as full recovery. The represented stress schedule lacks the contract needed for that test.

Source S1: changing forecast vintages and unresolved stress semantics. Croushore–Stark (2000); Gneiting–Raftery (2007). No historical forecast-accuracy estimate is available.

TECHNICAL FOUNDATION • VARIATION AND SELECTION

A revision sequence is not an operating process chart

[RC] The private audit computes 17 mechanical XmR screens. None is a commensurate observed operating history that supports a natural-process signal claim.

Represented sequence	Permissible interpretation
Plan paths	[AI] Deterministic input shapes. Moving-range limits describe arithmetic variation only; they cannot classify delivered performance.
Forecast vintages	[AI] Information, actuals and model assumptions change between observations. Crossings cannot diagnose a stable operating process or an intervention effect.
Overlapping averages	[AI] A trailing average and a year-to-date average reuse observations. They neither reveal the monthly sequence nor establish acceleration.
All-zero recording	[AI] Zero moving range is degenerate. Without validated observed zeros it does not mean perfect stability, reliability or zero benefit.

[PD] For an eligible ordered series, $MR(t) = |x(t) - x(t-1)|$, I limits = $\text{mean}(x) \pm 2.66 \text{ mean}(MR)$, and MR upper limit = $3.268 \text{ mean}(MR)$. Screen spacing, phase, dependence and denominators first. The present computations remain non-diagnostic. Missing counts preclude a valid Laney p-prime analysis; rounded rates cannot manufacture them.

[AI] No fixed worst-N treatment cohort or follow-up panel is supplied for this booking case. Underlying program targeting does not identify a cohort effect; field evaluation must freeze selection and quantify regression-to-the-mean sensitivity. Other selection includes favorable scenarios, recoverable files and completed decisions. Register every review and declared scenario, including unfavorable or unresolved cases.

[PD] Complexity, formula architecture, node mix and analyst capability are moderators. Preregister interactions and report uncertainty. Unchanged comparison-group cells delimit represented scope; they are not an untreated causal control or evidence of absent spillovers.

Source S1: no comparable actual panel; private mechanical screens. NIST individual-chart formulas are used without a process interpretation.

TECHNICAL FOUNDATION • CONTRIBUTION ATTRIBUTION

Model authorship does not establish entitlement or delivery

[AI] Process tracing must evaluate a specific contribution hypothesis against alternative explanations. Repetition within one evidence lineage does not create independent corroboration.

Contribution hypothesis	Current test and next evidence
Analysis and model editing	[HR/AI] Provisional registry-level hoop support: the register describes first-person booking and recalculation. Recover original versions and authenticated edit history to distinguish authoring from copied or assisted text.
Scenario communication	[HR/AI] Provisional registry-level hoop support for drafting. Original sent message, recipients and exact attached version are required to establish transmission and sequence.
Entitlement approval	[AI] Unestablished. Persistence in later saves and a request for explanation do not demonstrate authorization. Require a scoped approval record and accountable approver.
Field delivery	[AI] Unestablished. Planning cells and named owners do not document completed actions. Recover dated receipts, eligible units and independent verification.
Customer or cash impact	[AI] Unestablished. Obtain identified mature-outcome effects, a cost bridge and finance evidence. A positive model position is insufficient.

[PD] A hoop test documents a necessary condition, not sufficiency. A smoking gun must be unlikely under credible alternatives; a doubly decisive test must also exclude them. No present claim earns those stronger grades. Label the grade next to the claim and retain the timing, alternative and provenance in the private ledger.

[AI] Chronology remains unresolved because represented message and save times do not establish the claimed sequence. Do not infer which adjacent edits the case subject made or count a neighboring outcome bridge as part of this project’s result.

Source S1: registered role descriptions, chronology and versioned cells. Original authorship, approval, delivery and outcome records remain unauthenticated or unavailable.

TECHNICAL FOUNDATION • PROSPECTIVE EVALUATION

Randomize the review policy before measuring its decision cost

[PD] Pair independent planning teams on mature baseline burden and portfolio complexity; randomize one per pair to FCP and one to documented usual review.

[PD] Register eight weekly baseline decision cohorts at weeks –20 through –13 and eight post cohorts at weeks 0 through 7. Each outcome closes 90 days after its original scheduled decision. Finish baseline follow-up before assignment. Include every upstream-eligible slot, including unsubmitted, held and withdrawn cases; keep controls unchanged through the last outcome horizon.

Design component	Prespecified safeguard
Primary burden	[PD] Ordinary review, rework, corrections and nonduplicated unresolved/delay costs. Exclude incremental FCP costs booked separately. Justified negative investigations retain their resource cost without being labeled failure.
Independent adjudication	[PD] Use common sanitized packets, masked protocol labels and a frozen rubric. Incomplete reference truth stays unresolved; publish bounds and disagreement instead of forced certainty.
Quality gates	[PD] Unsupported approval, false holds, decision coverage, unresolved evidence and serious downstream harm. Predeclare margins and multiplicity; the burden test alone does not power them.
Spillovers and attrition	[PD] Separate shared reviewers/templates or randomize higher-level teams. Retain assigned slots after staff changes; collect reasons and bound missing outcomes.

[PD] Adjust control-minus-FCP follow-up pair means by .50 times baseline difference; fix this coefficient from an independent pilot. Cluster at the randomization unit. Inspect baseline trajectories, composition and fidelity; non-significant differences do not prove exchangeability. If randomization fails, redesign and justify a group-time comparison with anticipation and parallel-trend sensitivities; do not relabel uncontrolled before/after as this trial.

Proposed design, with synthetic power in the following model section. Vickers–Altman (2001) supports prespecified baseline adjustment; customer impact is separate and unpowered.

TECHNICAL FOUNDATION • INTELLECTUAL CONTRIBUTION

The protocol makes computational credit independently falsifiable

[PD] FCP is a proposed framework: type the contrast, replay coherent edits, crosswalk rates to units, authorize the decision and validate the mature outcome.

Proposition	Test and disconfirming observation
P1 • Typed attribution	[PD] On registered adversarial cases, FCP reduces unsupported booking attributions versus usual review. No reduction, or an exceeded false-hold margin, defeats the specified claim.
P2 • Replay completeness	[PD] Under one engine and a complete coherent edit partition, allocated edits reconcile within tolerance. Failure rejects completeness of that replay; success does not establish field causality.
P3 • Evidence-state validity	[PD] Typed zeros reduce false zero-effect conclusions on externally labeled records without materially degrading true-zero recognition or coverage. Test the full confusion matrix.
P4 • Net decision value	[PD] The lower confidence bound for burden reduction exceeds the identical break-even while quality gates pass. A spanning interval is inconclusive; an upper bound below break-even contradicts positive net value under the stated economics.

[AI] Sargent’s verification/validation distinction explains why replay can pass while field benefit remains unknown. Hermans and colleagues motivate dependency visibility. Model-attribution research shows why baseline, grouping and context determine allocation; the case extends that logic to credit governance, without empirically validating FCP.

[AI] Campbell’s measurement critique and strategic-measurement research motivate a separate check of incentives and booking-to-outcome links. They do not establish manipulation here. Process-chart discipline prevents target edits from being narrated as operating signals; learning requires correction ownership and outcome feedback beyond an amended spreadsheet.

Proposed propositions; primary literature is used to constrain interpretation, not certify performance. FCP has not been prospectively tested in this case.

TECHNICAL FOUNDATION • LIMITATIONS AND REPRODUCIBILITY

Release the interpretation with the unresolved evidence intact

[AI] This is an evidence-reconciliation and proposed-governance case. Its reproducible findings concern represented arithmetic and inferential boundaries.

Limit	Consequence and closure evidence
Originals and lineage	[HR] The register is available; original calculation files and transmission evidence are not authenticated. Recover originals, hashes, engine and external dependencies before claiming replay.
Measurement and timing	[AI] Scope, zero meanings, month alignment and scenario semantics are unresolved. Authenticate a metric/clock/version contract; retain source forks rather than averaging incompatible states.
Identification	[AI] No identified booking-only computational effect, intervention effect or historical forecast accuracy. Four-state replay and a separate field comparison answer different missing questions.
Contribution	[AI] Registry-level role evidence is provisional. It does not establish authorization, implementation or outcome ownership.
Economics and transport	[PD] Costs, realization, uncertainty and treatment effect are hypothetical. Pilot independent teams and price categories; customer impact and cash cannot be imported from model outputs.

[PD] Private reproduction includes the frozen extraction, source-cell ledger, hashes, exact arithmetic, alternate baselines, month vectors, formula-screen exclusions, synthetic parameter provenance and executable model checks. Rerun every headline after changes. Recovering originals adds evidence; it must not overwrite the preserved registered state.

[PD] Public files retain anonymous roles, protected source aliases and indexed or relative quantities. Raw values, original titles, links and identity crosswalks stay private. Inspect companion outputs for invertible anchors. The final author-led, agent-assisted referee report evaluates claim discipline and design completeness; it is not external peer review or impact certification.

Source S1 and the private audit establish the current evidence boundary. The proposed evaluation and economics remain conditional on future data and implementation.

PROPOSED DECISION MODEL · ECONOMICS

Price the governance effect without borrowing customer savings

[PD] All quantities in this section are synthetic planning assumptions. Historical intervention effects, avoided cost and released cash remain unestimated.

[PD] The estimand is the 90-day mean ordinary decision-resource burden reduction per fixed eligible slot over eight post weeks. One pretrial burden pool is indexed to 100. Let b be the absolute ITT reduction in those index points. Net value= $a \times q \times b - K$; base $a=.70$, $q=.90$ and $K=4$ give break-even $b=6.3492$.

Component	Definition preventing double counting		
a · realization	[PD] Share of released burden converted into usable capacity or expense reduction. This is not another avoidability correction; cash requires a documented release mechanism.		
q · unique credit	[PD] Share not credited elsewhere. It removes economic duplication, not a second causal-effect correction.		
K · new protocol cost	[PD] Incremental FCP training, setup and audit cost allocated to the same eight-post-week slot pool and 90-day resource horizon. Exclude from ordinary burden and subtract once.		

Benefit b	a=.40	a=.70	a=1.00
6	-1.84	-0.22	1.40
12	0.32	3.56	6.80
18	2.48	7.34	12.20

[PD] Grid fixes $q=.90$ and $K=4$. At $q=.70$ the base break-even rises to 8.1633 points. Prices and burden composition must remain compatible; if treatment changes the mix, estimate category-priced ITT directly. Delay penalties cannot duplicate labor or opportunity losses. Guardrails prevent a cheap all-hold policy from qualifying.

Synthetic model only. b is a fixed-reference index difference, not a percentage of changing control burden. Customer value and information value are excluded.

PROPOSED DECISION MODEL · FINITE-PAIR POWER

Make the trial precise enough to resolve the economic threshold

[PD] The planned comparison randomizes independent teams within pairs. Eight mature baseline cohorts precede assignment; eight post cohorts retain the original 90-day outcome clock.

[PD] Assume team-period Gaussian residuals with SD 24 and covariance $\sigma^2[\rho+(1-\rho)\phi^{|t-s|}]$, $\rho=.25$ and $\phi=.40$. Baseline weeks are $-20\dots-13$; post weeks $0\dots7$. Equal slot weights and independent teams are assumptions. Shared pair shocks cancel only if both teams are affected identically.

[PD] Define D as control-minus-FCP phase means. Test adjusted pair benefit $D_{\text{post}}-.50D_{\text{base}}$. The .50 slope is fixed from an assumed independent pilot, not estimated in this trial. Its variance is 349.5243; the one-sample Student t has $m-1$ degrees of freedom for m iid Gaussian pairs.

Pairs / teams	80% MDE	Economic power	Both gates
24 / 48	11.16470	58.4%	Fail
48 / 96	7.71891	88.1%	Fail
70 / 140	6.34910	96.8%	Pass
80 / 160	5.92834	98.3%	Pass

[PD] Mathematical minimum: 70 pairs/140 teams. The 80-pair candidate adds a buffer, conditional on pilot validation. MDE targets two-sided $\alpha=.05$, 80% power. Economic power is $P(\text{lower endpoint of two-sided 95\% CI} > 6.3492)$ at benefit 15. Team availability is unknown; these are conditional designs, not achieved power.

Synthetic finite-t power integrates the normal/ χ^2 scale mixture; numerical mass and quadrature convergence are checked. No current trial baseline or variance is observed.

PROPOSED DECISION MODEL · DESIGN SENSITIVITY

A pilot must validate covariance before the sample is committed

[PD] The fixed adjustment coefficient preserves the randomized mean effect, but a poor coefficient can increase variance. Matching does not guarantee precision.

Team SD	$\beta=0$	$\beta=.50$	$\beta=1.00$
18	58	41	51
24	102	70	89
30	158	109	138

[PD] Entries are minimum pairs meeting both break-even detectability and economic-confidence power; teams=2×pairs. Every cell retains $\rho=.25$, $\phi=.40$, benefit 15 and the same economic threshold. Freeze β from an independent pilot before randomization; trial estimation requires an ANCOVA leverage correction and different degrees of freedom.

[PD] For phase-difference covariance V , adjusted variance= $V_{\text{post}}+\beta^2V_{\text{base}}-2\beta\text{Cov}(\text{base},\text{post})$. Conditional Student power is exact under iid normal pair scores and constant mean effect; covariance values, normality and transport to eligible teams remain planning assumptions.

[PD] The 70-pair minimum is knife-edge: MDE 6.34910 against break-even 6.34921. An 80-pair buffer still needs a pilot. Re-power variance uncertainty, unequal slots, missing teams, portfolio changes and contamination. If teams are fewer, obtain supported variance reduction, a longer coherent design or a different randomization unit.

[PD] Prespecify unsupported-credit, false-hold, coverage and harm margins separately. They require their own sample-size and multiplicity assessment. Passing the primary burden gate alone is insufficient to authorize deployment; no customer-outcome claim follows from this governance trial.

Synthetic design sensitivity. A covariance pilot, eligibility census and guardrail sizing remain prerequisites; no undisclosed expansion of the project population is assumed.

PROPOSED DECISION MODEL · VALUE OF INFORMATION

Commission evidence only when it improves the next decision

[PD] A separate hypothetical model asks whether an evidence audit improves a release/hold/replay decision. It does not assume a positive customer effect.

Action / net utility	Valid	Incompatible	Unverified
Release credit	20	-50	-35
Conditional hold	0	-4	-8
Commission replay	8	10	-12

[PD] States use a primary-blocker hierarchy: incompatible basis takes precedence; otherwise required field evidence is unverified or all credit standards pass. Unverified includes missing implementation or attribution evidence. Priors are .50/.30/.20. Utilities already include action costs. Replay can repair model incompatibility but cannot establish a field effect.

Audit accuracy / cost	EVSI	Net EVSI
55% / 0.5	0.000	-0.500
75% / 1.0	1.675	0.675
90% / 2.0	4.580	2.580

[PD] Each signal identifies its state with the stated accuracy; remaining likelihood splits equally across the other two labels. Without information the best action is commission replay, utility 4.60. EVPI=6.80. EVSI averages the best posterior action value minus the prior optimum; net EVSI subtracts audit cost separately.

[PD] Audit cost is separate from the subsequent replay action cost already in utilities. Low/base/high are assumptions, not confidence bounds. Validate priors, likelihoods and payoffs. Never add this value to prevented review loss when they describe the same consequences; no information value is claimed historically.

Synthetic Bayesian decision model. Utility table, likelihoods, posterior actions, EVPI bound and arithmetic are reproduced privately and in anonymous public calculations.

Read forecast credit as a typed claim with a verification burden

[AI] The case connects operational commitments, analytical models and implementation evidence through separate approval gates. The observed model arithmetic does not bypass those gates.

Claim class	Meaning in this case
[HR] Historical record	A statement or represented artifact in the supplied project register. Original workbooks, sent communications and field results were not independently authenticated.
[RC] Recalculated	Arithmetic from registered inputs or a declared synthetic design. Reproduction retains the assumptions and evidence class of those inputs.
[AI] Analytical inference	A constrained interpretation or implication. It is neither an identified historical effect nor proof of an implemented protocol.
[PD] Proposed design	A future credit protocol, evidence contract, evaluation or synthetic planning model. Its parameters are not measured project results.

[AI] The positive cushion is 2.71% of the whole-save upward revision. The later-month booking differs from its narrative, and another row’s blanket edit description fails arithmetic reconciliation. These findings change how credit should be reviewed; they do not identify the intervention’s delivered benefit.

[AI] FCP separates computational credit, authorization, delivery and outcome validation. Spreadsheet research supports dependency inspection; model-validation research distinguishes verification from operational validity. Forecast literature requires matched targets, and causal methods require a treatment comparison. The case applies these distinctions without claiming to have tested FCP.

[PD] Reproduction chain: frozen register and hashes → source-cell ledger → exact revision and version calculations → protected exhibits → separately labeled synthetic evaluation → independent numerical and artifact review. Preserve missing replay states and alternate source interpretations rather than filling them with inferred results.

DISTRIBUTION

[PD] Share this anonymized case PDF. Keep the separate evidence annex and reproduction package private: they contain original identities, source locations, exact dates and raw quantities.

S1: supplied project register. Historical quantities are indexed or relative. Proposed study assumptions, economic inputs and information models are synthetic and explicitly labeled.

REFERENCES / PRIMARY METHODS AND SUBSTANTIVE LITERATURE

Verify the model before interpreting its allocation

Sources support the interpretation and proposed design. They do not authenticate the project, validate its forecasts or establish its effects.

Hermans, F., Pinzger, M., & van Deursen, A. (2011). Supporting Professional Spreadsheet Users by Generating Levelled Dataflow Diagrams. ICSE, 451–460. Opened author full text. Dependency visibility supports comprehension; adjacent labels do not establish a computational link.

Sargent, R. G. (2020). Verification and Validation of Simulation Models: An Advanced Tutorial. Winter Simulation Conference, 16–29. Official full paper. Verification, data validity and operational validity address different intended-use questions.

Sundararajan, M., & Najmi, A. (2020). The Many Shapley Values for Model Explanation. PMLR, 119, 9269–9278. Original full text. Baseline, grouping and context affect model attribution; the replay allocation here remains a declared convention.

Ma, S., & Tourani, R. (2020). Predictive and Causal Implications of Using Shapley Value for Model Interpretation. PMLR, 127, 23–38. Original full text. Model explanation and causal interpretation require different assumptions.

W3C (2013). PROV Primer. Official provenance guidance. Versioned entities, activities and agents make lineage explicit. Recorded provenance does not establish numerical correctness, authorship authenticity or outcome validity.

Croushore, D., & Stark, T. (2000). A Real-Time Data Set for Macroeconomists: Does Data Vintage Matter for Forecasting? Federal Reserve Bank of Philadelphia, WP00-6. Official text. Preserve issue-time data and distinguish revisions from forecast-target realizations.

Gneiting, T., & Raftery, A. E. (2007). Strictly Proper Scoring Rules, Prediction, and Estimation. JASA, 102, 359–378. Author full text. Proper scoring applies to declared predictive distributions; two stress scenarios do not imply calibrated probabilities.

Original research, official guidance and author manuscripts. Access and interpretive limits are recorded in the private methods memorandum. No source supplies a historical effect for this project.

REFERENCES / PRIMARY METHODS AND SUBSTANTIVE LITERATURE

Measurement must preserve the meaning of a credit claim

Sources support the interpretation and proposed design. They do not authenticate the project, validate its forecasts or establish its effects.

Wang, R. Y., & Strong, D. M. (1996). Beyond Accuracy: What Data Quality Means to Data Consumers. *Journal of Management Information Systems*, 12(4), 5–33. Author-hosted text. Fitness for a decision includes context, representation and accessibility.

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20, 37–46. Original methodological article. Chance-corrected agreement supports taxonomy auditing; agreement does not establish ground truth.

Campbell, D. T. (1976). Assessing the Impact of Planned Social Change. Original occasional-paper record, ERIC. Evaluation pressure can alter indicator behavior. This case identifies a governance risk, not demonstrated manipulation.

Ittner, C. D., Larcker, D. F., & Randall, T. (2003). Performance Implications of Strategic Performance Measurement in Financial Services Firms. *Accounting, Organizations and Society*. Author institutional summary. Selected measures need validated links to strategic results.

NIST/SEMATECH. e-Handbook of Statistical Methods: Individuals Control Charts. Official formula guidance. Mechanical limits do not create a process interpretation for deterministic plans, changing vintages or unverified zero cells.

The W. Edwards Deming Institute. The Funnel Experiment. Institutional account of tampering with a stable process. Supports checking whether variation justifies an action trigger; no stable historical operating panel is established here.

Tucker, A. L., & Edmondson, A. C. (2003). Why Hospitals Don't Learn from Failures. *California Management Review*. Author working paper. Correcting an immediate record differs from changing the organizational causes and learning process.

OECD / European Commission JRC (2008). Handbook on Constructing Composite Indicators: Methodology and User Guide. Institutional guidance. The proposed burden measure needs explicit components, frozen prices, non-overlap and sensitivity; a weighted total cannot replace its guardrails.

Original research, official guidance and author manuscripts. Access and interpretive limits are recorded in the private methods memorandum. No source supplies a historical effect for this project.

REFERENCES / PRIMARY METHODS AND SUBSTANTIVE LITERATURE

Evaluate the policy without importing the intervention's benefit

Sources support the interpretation and proposed design. They do not authenticate the project, validate its forecasts or establish its effects.

Vickers, A. J., & Altman, D. G. (2001). Analysing Controlled Trials with Baseline and Follow Up Measurements. *BMJ*, 323, 1123–1124. Primary text located; direct access limited. Supports prespecified baseline adjustment. The paired clustered model here is a separate planning derivation.

Hernán, M. A., & Robins, J. M. Causal Inference: What If. Author-hosted book and current materials. Treatment definitions, assignment-based estimands and selection. Eligibility must precede any protocol-induced submission, holding or withdrawal decision.

Collier, D. (2011). Understanding Process Tracing. *PS: Political Science & Politics*, 44, 823–830. Necessary and discriminating evidence for a specific contribution hypothesis. Repeated files from one lineage are not independent corroboration.

Callaway, B., & Sant'Anna, P. H. C. (2021). Difference-in-Differences with Multiple Time Periods. *Journal of Econometrics*. Author manuscript. An observational fallback needs a suitable group-time estimand and explicit identification assumptions; it does not inherit randomized-trial power.

Roth, J. (2022). Pretest with Caution: Event-Study Estimates after Testing for Parallel Trends. *American Economic Review: Insights*, 4, 305–322. A nonsignificant pretrend test does not establish a valid observational comparison.

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*. EVSI values improved decisions under a stated information model, without presuming a positive field effect.

Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice*, 3rd ed., §1.2, "Forecasting, goals and planning." Official author text. A desired outcome, a prediction and an action plan require different interpretation and evaluation.

Original research, official guidance and author manuscripts. Access and interpretive limits are recorded in the private methods memorandum. No source supplies a historical effect for this project.

FINAL REFEREE REPORT / INTERNAL, AGENT-ASSISTED REVIEW

The case separates modeled recovery from verified performance

[AI] The author and parallel agents challenged source fidelity, analytical methods and executive use. This is internal review, not external academic or client certification.

Major issue	Resolution or remaining gate
Revision attribution	The whole-save change and scenario spread remain bundled contrasts. Missing replay states are uncomputed. The corrected period vector replaces an inconsistent verbal edit account.
Performance and zeros	Later forecasts are not realized forecast targets. All-zero cells retain unresolved evidence states. Neither persistence nor a missing attribution entry is treated as a field effect.
Source and contribution	Originals remain unauthenticated. Modeling and drafting receive provisional registry-level hoop support; approval, delivery and outcomes remain unestablished.
Evaluation and economics	Independent covariance, finite-t, exact-information and simulation checks pass 84 tests. External baseline adjustment, matched cost horizon and an explicit precision buffer constrain the design; guardrail sizing remains a prerequisite.

Source corrections carried into the case

[AI] The 2.71% cushion uses the whole-save revision denominator. Narrative and cell increments remain unresolved. Full attainment in one month does not automatically erase earlier cumulative deficits. Overlapping averages do not establish a monthly acceleration pattern.

Partner critique: would the COO act on page 1?

[AI] Yes: keep the scenarios visible, appoint a credit owner and require the controlled replay. The thin cushion and bundled edits make the verification decision concrete. The proposed gates connect analytics with accountable operating evidence while preserving the case’s planning boundary.

REVIEW DISPOSITION

[AI] Suitable as an anonymized evidence-reconciliation and proposed-governance case after numerical and artifact review. Historical customer impact, realized financial value and the feasibility of the proposed study remain unestablished.

Internal author-led review and independent agent challenge. Numerical reproduction, source-boundary checks and page-by-page artifact reviews are retained in the private package.