

Original Article

Designing Multi-Type Purchase Order Workflows through Stakeholder Synthesis: A Framework for Requirements Integration in Retail Supply Chains

***Aakanksha Polepalli**

Product Manager, Grocery Outlet, Emerville, CA, USA.

Abstract:

large retail supply chains increasingly rely on heterogeneous purchase-order (PO) workflows spanning routine replenishment, promotion, import, drop-ship, cross-dock, direct-store-delivery, and exception purchasing. These workflows cross merchandising, procurement, planning, logistics, finance, supplier operations, and information-technology boundaries, creating a requirements-integration problem that isolated functional elicitation cannot reliably resolve. Existing requirements-engineering and supply-chain integration studies associate stakeholder participation and information integration with superior performance, but these associations do not establish whether stakeholder synthesis itself causes operational improvement. This article develops a primary causal research design estimating the effect of Stakeholder-Synthesized Requirements Integration (SSRI) on workflow-induced PO exceptions. The proposed empirical setting exploits phased adoption of mandatory SSRI governance across procurement operating cells and uses a staggered difference-in-differences design with cohort-time average treatment effects. SSRI requires cross-functional stakeholder coverage, PO-type decomposition, requirement reconciliation, traceability, scenario validation, and joint sign-off. The primary outcome is workflow-induced exceptions per 100 eligible POs; secondary outcomes include first-pass completion, cycle time, manual handoffs, post-release corrections, and requirements-defect escape. The design pre-specifies modern event-study diagnostics, falsification tests, co-intervention restrictions, and sensitivity analysis. Mechanisms operate through requirements completeness, rule coherence, reduced defect escape, and reduced manual rework, with stronger effects predicted under greater workflow complexity, supplier heterogeneity, demand volatility, and system fragmentation. Because retailer microdata were not supplied, no empirical coefficient is fabricated; instead, the paper provides an executable causal protocol and publication-ready results architecture.

Keywords:

Requirements Engineering, Purchase Orders, Stakeholder Synthesis, Retail Supply Chain, Causal Inference, Staggered Difference-In-Differences, Procurement Workflow, Enterprise Systems.

Article History:

Received: 04.01.2025

Revised: 09.02.2025

Accepted: 19.02.2025

Published: 28.02.2025

1. Introduction and Causal Puzzle

Purchase orders constitute one of the principal transactional mechanisms connecting retail demand, sourcing, suppliers, transportation, inventory, receiving, and financial settlement. Although enterprise platforms often represent a purchase order as a standardized business object, the processes governing purchase orders are heterogeneous. Routine replenishment orders may be generated algorithmically and released with little human intervention. Promotional and seasonal orders can require budget controls, event-specific quantities, delivery windows, and rapid amendments. Import orders introduce longer lead times, documentation,



transportation milestones, and country-specific constraints. Drop-ship orders bypass conventional distribution-center receipt, while cross-dock and direct-store-delivery arrangements alter fulfillment, receiving, and settlement logic. A large retailer therefore operates not one PO workflow but a portfolio of related workflow types that reuse data, rules, suppliers, approval authorities, and technology components.

The practical difficulty is that knowledge required to design these workflows is distributed across organizational functions. Merchandising understands assortment and commercial intent; replenishment teams understand forecast and allocation policies; procurement understands sourcing and supplier agreements; logistics understands shipment, appointment, and receiving constraints; finance and accounts payable govern approval and settlement; supplier operations understand vendor capabilities; internal control functions establish authorization boundaries; and engineering teams translate the combined requirements into executable rules, services, events, interfaces, and validations. Each function can provide locally correct requirements while the end-to-end workflow remains globally inconsistent.

Consider an import promotion order. Merchandising may permit quantity changes until a commercial cutoff, procurement may permit supplier acknowledgement changes, logistics may require container booking before that same cutoff, and accounts payable may depend on a stable matching quantity. If these requirements are elicited independently, no individual statement is necessarily wrong, yet their combination can create a workflow that permits a state transition that downstream operations cannot support. Similar failures arise when a standard PO approval rule is reused for a promotional subtype, when drop-ship orders enter warehouse-oriented receipt paths, when supplier data judged complete upstream are insufficient for downstream settlement, or when cancellation logic handles an original order but not a partially acknowledged or split order. Such interface failures generate manual overrides, repeated approvals, cancel-and-recreate behavior, support tickets, delayed releases, and downstream matching errors.

Requirements-engineering scholarship has long emphasized requirements quality, stakeholder participation, communication, traceability, validation, and conflict resolution (Hofmann & Lehner, 2001; Damian & Chisan, 2006; Méndez Fernández et al., 2017). Stakeholder-identification research similarly demonstrates that deciding whose knowledge enters the requirements process is itself a substantive problem (Sharp et al., 1999; Lim et al., 2010). Supply-chain research, meanwhile, associates information integration and coordination with operational performance (Vickery et al., 2003; Flynn et al., 2010; Prajogo & Lager, 2012), while e-procurement research links system and information quality to procurement outcomes (Vaidyanathan & Devaraj, 2008; Tai et al., 2010). These literatures jointly make stakeholder synthesis theoretically plausible, but they do not by themselves establish that introducing a stakeholder-synthesis intervention causes PO-workflow performance to improve.

The causal ambiguity is substantial. Suppose units reporting extensive stakeholder participation also have fewer workflow exceptions. First, stakeholder synthesis may genuinely improve requirements and reduce defects. Second, better-managed units may simultaneously use stronger requirements practices and deliver better operational performance, creating omitted-variable bias from managerial capability. Third, reverse causality may occur: units suffering unusually high exception rates may receive intensive cross-functional workshops precisely because performance is poor. Fourth, complex units may both require more stakeholders and produce more exceptions, confounding treatment and outcome. Fifth, the intervention may coincide with a technology migration, supplier onboarding initiative, training program, master-data cleanup, or additional implementation resources. A standard multivariable regression cannot make these alternatives disappear merely by adding observable controls.

The conventional observational model can be written as:

$$Y_{it} = \alpha + \beta \text{SSRI}_{it} + \gamma'X_{it} + \varepsilon_{it} \quad (1)$$

Where Y_{it} denotes workflow performance for operating cell i in period t , SSRI_{it} denotes exposure to stakeholder-synthesized requirements integration, X_{it} contains measured covariates, and the error term captures remaining determinants. A causal interpretation of beta requires that treatment exposure be conditionally independent of the unobserved determinants of the outcome. If managers prioritize troubled cells, if high-capability teams adopt earlier, or if coincident transformations are incompletely measured, this condition fails. Correlation is therefore not the target estimand.

The missing object is the counterfactual. Let $Y_{it}(1)$ denote the outcome cell i would experience at time t under SSRI and $Y_{it}(0)$ the outcome it would experience without SSRI. The individual causal effect is

$$\tau_{it} = Y_{it}(1) - Y_{it}(0) \quad (2)$$

After adoption, the treated cell reveals $Y_{it}(1)$, but $Y_{it}(0)$ is unobservable. A before-after comparison cannot recover it because retail seasonality, demand conditions, supplier mix, enterprise releases, and organizational learning also change through time. A contemporaneous treated-versus-untreated comparison is likewise inadequate if treatment timing is systematically related to baseline capability or operational risk. The research design must therefore construct a credible untreated trajectory for treated operating cells.

This paper defines the treatment as Stakeholder-Synthesized Requirements Integration (SSRI), a mandatory governance protocol for the design or modification of multi-type PO workflows. A cell is treated when requirements changes must satisfy six elements: cross-functional stakeholder coverage; explicit decomposition by PO type; reconciliation of conflicting rules and terminology; traceability from stakeholder need to executable rule and acceptance criterion; scenario validation covering normal, boundary, amendment, cancellation, and exception paths; and documented cross-functional sign-off. SSRI is not equivalent to generic digitization, a software upgrade, more meetings, conventional user participation, or UAT alone. The treatment is a structured requirements-integration regime.

The primary outcome is the weekly rate of workflow-induced exceptions per 100 eligible POs. Included events are incorrect routing, invalid state transitions, requirements-related missing data, incompatible PO-type processing, rule conflicts, avoidable manual overrides, workflow-driven cancel-and-recreate behavior, and requirements/configuration failures. External supplier production failures, weather disruptions, and carrier delays after a correctly transmitted PO are excluded because they lie outside the proposed causal mechanism. Secondary outcomes include first-pass completion, creation-to-release cycle time, manual handoffs, repeated state transitions, post-release corrections, and requirements-defect escape.

The causal research question is: What is the causal effect of mandatory SSRI on workflow-induced PO exception rates, through what mechanisms does the effect operate, and under what levels of workflow complexity, supplier heterogeneity, demand volatility, and system fragmentation is the effect stronger or weaker?

The paper contributes in four ways. First, it converts stakeholder synthesis from an abstract perception into an auditable organizational treatment. Second, it introduces an explicit counterfactual strategy based on phased SSRI adoption. Third, it links requirements-level mechanisms to transaction-level outcomes. Fourth, it treats effect heterogeneity as theoretically meaningful rather than as exploratory segmentation. The empirical design uses modern staggered difference-in-differences rather than a conventional two-way fixed-effects coefficient because staggered timing and heterogeneous treatment effects can produce misleading comparisons (de Chaisemartin & D'Haultfœuille, 2020; Goodman-Bacon, 2021). The study cutoff is February 28, 2025. Because no retailer microdata were supplied with this manuscript, no coefficient, standard error, p-value, sample size, or treatment-effect magnitude is fabricated. The contribution is a publication-ready causal design that specifies exactly what must be estimated before causal language is warranted.

2. Theory, Mechanisms, and Causal Hypotheses

Organizational information-processing theory provides the principal causal foundation. Galbraith (1974) argues that uncertainty increases an organization's information-processing requirements, while Tushman and Nadler (1978) conceptualize organizational design as the alignment of information-processing capacity with task demands. Daft and Lengel (1986) further distinguish uncertainty from equivocality: organizations must not merely acquire more information but also resolve conflicting interpretations. Multi-type PO design contains both problems. Designers cannot anticipate every future combination of supplier capability, PO subtype, amendment, approval, fulfillment route, and downstream response, and separate functions can attach different meanings to the same workflow concepts.

SSRI is therefore theorized to change outcomes by increasing the organization's capacity to reconcile distributed knowledge before it becomes executable workflow logic. The central distinction is between stakeholder presence and stakeholder synthesis. A project may include every relevant function and still produce fragmented requirements if each function submits independent requirements, conflicting terms remain unresolved, and cross-type dependencies are not validated. Synthesis requires elicitation plus conflict resolution, traceability, cross-type reasoning, and joint scenario validation.

The first mechanism is requirements completeness. Let R^* represent the complete set of workflow-relevant requirements and R_j the subset known by stakeholder group j . Independent elicitation attempts to approximate R^* by the union of stakeholder-specific sets, yet

a simple union cannot guarantee that critical requirements were elicited, preserved through handoffs, or made logically compatible. SSRI can be represented as a synthesis operator $S(.)$ applied to the stakeholder requirement sets:

$$\hat{R}_{SSRI} = S(R_1, R_2, R_i) \quad (3)$$

The operator denotes terminology normalization, conflict detection, dependency analysis, priority decisions, PO-type differentiation, and traceability. The expected consequence is not necessarily more requirements. It is fewer consequential omissions and contradictions. Accordingly, H1 states that mandatory SSRI causally reduces workflow-induced PO exception rates relative to the counterfactual outcome for treated operating cells had they not yet adopted SSRI.

The second mechanism is rule coherence. A PO workflow can be represented as a state-transition system in which an action from state k to state l is valid only for particular PO characteristics x and PO type τ . The probability that a transition is permissible is therefore conditional on both transaction attributes and PO type. Requirements failures occur when a workflow implicitly assumes that a rule is invariant across PO types even though it is not. Domestic and import orders can share an approval state yet require different release conditions; warehouse and drop-ship orders can share item data yet require different receiving behavior. By requiring explicit PO-type decomposition, SSRI should expose such differences before implementation. H2 therefore predicts that SSRI causally increases first-pass workflow completion and reduces routing-related manual intervention.

The third mechanism concerns requirements-defect escape. A design defect becomes operationally costly only if it survives elicitation, specification, review, testing, and validation. Define the requirements-escape rate as the share of requirements defects first discovered after production activation. Cross-functional scenario validation should move defect discovery earlier in the lifecycle. H3 therefore predicts that SSRI causally reduces the proportion of requirements defects first discovered in production.

The fourth mechanism is manual rework. Incomplete or contradictory rules frequently convert predetermined workflow decisions into human exceptions. Employees investigate, reroute, override, reopen, cancel, recreate, or manually correct transactions. Let H_{it} represent manual handoffs per eligible PO and W_{it} repeated transitions to a prior workflow state. The theory predicts negative treatment effects on both measures. H4 states that SSRI causally reduces manual handoffs and workflow rework, thereby improving processing efficiency and reducing the opportunity for subsequent transactional errors.

The mechanisms are ordered temporally. The intervention first changes how requirements are produced, then changes the quality and coherence of released workflow rules, and only after deployment can transaction outcomes respond.

SSRI $\rightarrow$ Requirements synthesis $\rightarrow$ Validated specification $\rightarrow$ Workflow implementation $\rightarrow$ Operational outcomes (4)

This sequence generates a falsifiable dynamic prediction. H5 states that operational treatment effects should not systematically precede SSRI adoption; effects should emerge only after synthesized requirements can influence production behavior and should persist if the redesigned workflow remains operative. Pre-treatment improvements of similar size to post-treatment improvements would be inconsistent with the proposed causal story unless teams demonstrably changed behavior in anticipation of formal adoption.

The value of SSRI should also depend on the information-processing burden of the workflow. The first moderator is baseline workflow complexity. A complexity index can combine the number of workflow branches, organizational roles, technical interfaces, PO-type variants, and explicit exception paths. High complexity increases the number of potential hidden dependencies and contradictory rules. H6 predicts a larger absolute reduction in exceptions among cells with greater pre-treatment workflow complexity.

Supplier heterogeneity is the second moderator. Suppliers differ in geography, lead time, EDI or API capability, acknowledgement processes, fulfillment arrangements, packaging conventions, payment terms, and direct-to-consumer capabilities. Greater heterogeneity expands the number of conditions the workflow must represent. H7 predicts stronger beneficial effects of SSRI where supplier heterogeneity is high.

The third moderator is technical fragmentation. Purchase orders may traverse ERP, procurement, integration middleware, supplier networks, transportation systems, warehouse platforms, and financial applications. Each interface creates semantic and technical dependencies. H8 predicts larger benefits where baseline system fragmentation is high, provided that fragmentation does not

become so technically binding that requirements improvements cannot be implemented. The relationship may therefore be nonlinear: stakeholder synthesis can identify cross-system failures without necessarily eliminating architectural constraints.

Demand volatility is the fourth moderator. Volatile demand increases amendments, expedited releases, quantity changes, allocation shifts, and cancellations, thereby exercising exception pathways more frequently. H9 predicts larger SSRI effects on amendment- and routing-related exceptions under greater pre-treatment demand volatility.

A strong causal theory also specifies outcomes that should not respond. If SSRI operates through workflow requirements, it should not mechanically improve an external supplier's production reliability or reduce a carrier delay that occurs after a correct order message has been transmitted. H10 therefore predicts approximately null effects on pre-specified negative-control outcomes outside the workflow's causal reach.

The causal structure can be summarized as follows. Baseline managerial capability, organizational complexity, supplier composition, and historical operational performance may influence treatment timing and also influence outcomes; they are potential confounders. SSRI affects requirements completeness and rule coherence, which affect defect escape and manual rework, ultimately affecting workflow outcomes. Calendar and demand shocks affect outcomes and may also affect treatment timing if rollout decisions react to them. Post-treatment mechanisms are therefore not ordinary covariates in the primary total-effect specification. Conditioning on them could block part of the treatment effect or induce post-treatment bias.

Several competing pathways require explicit empirical attention. First, platform modernization can coincide with SSRI and independently improve outcomes. Second, master-data cleansing can reduce exceptions without changing requirements governance. Third, implementation staffing and training can improve performance through attention rather than rule quality. Fourth, selecting high-error cells can create regression to the mean. Fifth, supplier composition changes can mimic a workflow improvement. The identification strategy consequently treats theory as a set of restrictions on what patterns should count as causal evidence: credible pre-treatment comparability, no anticipatory effects, mechanism-proximal changes in predicted directions, and approximately null effects on outcomes outside SSRI's plausible reach.

3. Identification Strategy and Data

The principal methodology is staggered difference-in-differences exploiting phased adoption of SSRI across procurement operating cells. This design is appropriate when an enterprise cannot implement the governance protocol everywhere simultaneously because facilitation, change capacity, or release windows are limited. The design is not causal merely because adoption is staggered. Its credibility depends on why units enter treatment when they do. The analysis therefore combines modern staggered-treatment estimators with a documentary audit of rollout timing.

The preferred treatment-assignment unit is the narrowest organizational cell at which SSRI is jointly mandated and operationally meaningful. Depending on the enterprise, this may be a region by business domain by PO-type family. If the governance mandate is assigned at a higher level, such as a region or division that higher level becomes both the treatment unit and the relevant clustering level. Measurement is aggregated to cell-week. Weekly observations preserve treatment timing and retail seasonality while providing sufficient transaction volume for stable rates.

Let G_i denote the first week in which operating cell i is subject to mandatory SSRI governance. Treatment is absorbing:

$$D_{it} = 1(t \geq G_i) \quad (5)$$

The analysis preserves three dates: the governance-mandate date, the first completed SSRI requirements gate, and the first production activation of an SSRI-designed workflow. The mandate date is primary because it creates an intention-to-treat-like estimand. Defining treatment by successful compliance would be problematic because high-capability cells may comply faster, making compliance itself endogenous. Production activation is used to interpret operational lags and as an alternative event-time origin.

The empirical design requires genuine enterprise records. PO transaction data should provide identifiers, timestamps, PO type, supplier, region, business unit, quantities, values, amendments, cancellations, releases, and status history. Workflow-event logs should identify states, transitions, approvals, validation failures, retries, manual actions, reassignments, and exception codes. Requirements

repositories should provide requirement IDs, stakeholder source, PO-type applicability, revisions, traceability, acceptance criteria, conflict-resolution history, and links to defects where available. Governance records should independently establish treatment dates from rollout calendars, workshop assignments, signed decision matrices, approvals, and release records. Incident systems should supply root cause and discovery stage. Supplier master data should capture technological capability, location, fulfillment mode, tenure, and transaction volume. Contextual data should identify promotions, outages, major supplier changes, platform releases, and other enterprise initiatives.

The ideal data lineage is Requirement -> Workflow Rule -> Release -> Operating Cell -> PO -> Workflow Event -> Exception. This lineage permits direct connection between a requirements-governance intervention and transaction-level consequences rather than reliance on retrospective perceptions.

The primary panel includes cells that are eligible for SSRI, possess consistently measured outcomes, have identifiable treatment timing, and have sufficient pre-treatment support. The study ends on February 28, 2025. The start date is not invented; it is the earliest point at which treatment and outcome measurement are consistently comparable. Preferably, major cohorts should have at least 52 pre-treatment weeks, with longer history used where available for trend diagnostics. Cells undergoing inseparable major ERP migration or equivalent transformation at the treatment date should be flagged for exclusion from the clean-rollout specification.

The principal outcome is the weekly workflow-induced exception rate:

$$\text{Exception Rate}_{it} = 100 \times (\text{Workflow-induced exceptions}_{it} / \text{Eligible POs}_{it}) \quad (6)$$

The exception ontology must be frozen before treatment-effect estimation. Included events are routing, validation, state-transition, requirements-data, configuration, and PO-type incompatibility failures. External supply and transportation disruptions are excluded. Secondary outcomes include median creation-to-release time, first-pass completion, and manual handoffs per PO, repeated workflow transitions, post-release correction rate, and requirements-defect escape. Ticket-derived classifications should be validated against transaction logs because SSRI may improve documentation and thereby change detection rather than true failure incidence.

Potential outcomes are indexed by treatment cohort. Let $Y_{it}(g)$ be the outcome if cell i first adopts in week g and $Y_{it}(o)$ the outcome if it remains untreated through week t . The target group-time estimand is

$$\text{ATT}(g,t) = E[Y_{it}(g) - Y_{it}(o) \mid G_i = g] \quad (7)$$

For t greater than or equal to g . Dynamic effects are indexed by event time $e = t - g$ and aggregated across supported cohorts. The headline estimand is the average treatment effect on treated operating cells, not a cross-sectional average effect for all theoretically eligible retailers.

The central identification assumption is conditional parallel trends. Conditional on pre-treatment covariates X_i , treated cohort g and its eligible not-yet-treated comparison cells must have experienced similar changes in $Y(o)$ in the absence of treatment. The assumption concerns changes rather than levels. A complex import cell can have a permanently higher exception level than a domestic replenishment cell and still provide useful counterfactual information if their untreated trends are comparable after appropriate conditioning and design restrictions.

Pre-treatment covariates include lagged outcome levels and slopes, PO volume, PO-type mix, supplier count, supplier heterogeneity, international-supplier share, workflow branch count, baseline manual-touch rate, number of connected systems, region, business domain, EDI/API penetration, historical demand volatility, and promotion exposure. Covariates potentially affected by SSRI after treatment must not be inserted as conventional controls in the total-effect model.

The primary estimator is the cohort-time ATT framework of Callaway and Sant'Anna (2021), preferably using a doubly robust implementation that combines an outcome-regression model with a generalized propensity model based only on pre-treatment information. Sant'Anna and Zhao (2020) motivate doubly robust DiD estimators. Conventional TWFE is not the headline estimator. Goodman-Bacon (2021) shows that staggered TWFE decomposes into multiple two-by-two comparisons, and de Chaisemartin and D'Haultfœuille (2020) demonstrate that heterogeneous effects can make standard coefficients difficult to interpret. TWFE can be reported only as a historical benchmark.

Dynamic treatment effects are estimated with a heterogeneity-robust event study. The event-study graph is not decorative; it tests whether apparent effects precede treatment, reveals implementation delay, distinguishes temporary disruption from persistence, and identifies treatment-effect evolution. Sun and Abraham (2021) provide a pre-specified robustness estimator for staggered event studies, while Borusyak et al. (2024) provide an imputation-based approach. Results should include both all-supported-cohort event time and a balanced-horizon version so long-run estimates are not driven solely by early cohorts.

No anticipation is required. If employees know in advance that SSRI is coming and change behavior before the recorded mandate, the treatment date must be moved to the first event capable of altering behavior, such as formal assignment or announcement. A transition window can also be excluded in sensitivity analysis. Treatment consistency requires that once a cell is classified as treated, the governance regime remains active. If treatment reverses, the baseline absorbing-treatment design is inappropriate for that cell.

The most important design-stage investigation is treatment-timing selection. Researchers should retrieve the earliest rollout plan and compare planned and actual treatment dates. A discrete-time treatment-hazard analysis should test whether recent exception spikes or other outcome deterioration predict immediate adoption after accounting for documented scheduling factors. If executives systematically accelerated troubled cells into treatment in response to information not captured in the dataset, the DiD interpretation weakens substantially. Conversely, timing driven mainly by recommitted administrative waves, facilitator availability, or capacity constraints is more compatible with identification.

Parallel trends cannot be proven by statistically insignificant leads. Roth (2022) shows why pretesting must be interpreted carefully. The study therefore uses visual lead estimates, joint tests, substantive equivalence bounds, and the sensitivity framework of Rambachan and Roth (2023). The latter quantifies how large a deviation from parallel trends would be required to overturn the substantive conclusion. A paper should report the conventional interval, the allowed deviation restriction, the robust interval, and the breakdown value rather than simply declaring that the assumption "holds."

Inference must respect assignment. Errors will typically be serially correlated within operating cells, making PO-level standard errors inappropriate. Bertrand et al. (2004) demonstrate the danger of ignoring serial correlation in DiD settings. The baseline analysis clusters at the treatment-assignment unit. If rollout occurs by region, region-level clustering is required. When the number of clusters is small, wild-cluster bootstrap inference should supplement asymptotic cluster-robust intervals.

Heterogeneous treatment effects are evaluated using moderators measured before treatment. Cells can be stratified using pre-specified terciles of workflow complexity, supplier heterogeneity, system fragmentation, and demand volatility. The relevant statistic is the difference between subgroup ATTs, not the comparison of whether one subgroup happens to be statistically significant and another does not. Continuous specifications may supplement but not replace these pre-specified contrasts.

The design pre-specifies multiple falsification tests. First, pseudo-treatment dates placed entirely in the genuine pre-treatment period should not generate systematic treatment effects. Second, negative-control outcomes such as exogenous supplier production shutdowns or carrier delays after correct transmission should remain approximately unchanged. Third, where some PO types are explicitly outside an SSRI release, excluded types provide a scope falsification test. Fourth, a clean-rollout sample removes cohorts overlapping major ERP migration, supplier-network conversion, master-data remediation, workforce restructuring, or unrelated automation. Fifth, stable-supplier and stable-PO-type samples address transaction-composition changes.

Missing cell-weeks must not be coded automatically as zero exceptions. Exclusions for insufficient history, treatment ambiguity, platform migration, missing workflow events, or severe spillover should be disclosed, and external validity should be limited primarily to treated cells and comparable retail environments.

4. Causal Results and Robustness: Pre-Specified Empirical Protocol

No retailer transaction-level dataset accompanied this manuscript. It would therefore be methodologically improper to report a synthetic sample size, fabricated coefficient, invented confidence interval, or simulated p-value as though it were an empirical finding. The results section instead specifies the evidence required for a defensible causal conclusion. When source data are analyzed, each table and figure described below can be populated directly from the pre-specified pipeline.

The empirical section should begin with treatment structure rather than regression results. Table 1 should report the number and definition of treatment-assignment units, treatment cohorts, never-treated units if any, eligible PO volume, PO-type distribution, pre- and post-treatment support, excluded co-intervention cells, and missing-data exclusions. A rollout figure should place operating cells on the vertical axis and calendar time on the horizontal axis, marking SSRI adoption and major concurrent transformations. This visual reveals whether adoption is genuinely staggered and whether treatment clusters around enterprise events.

The first causal figure should be a heterogeneity-robust event study. Event time is shown on the horizontal axis and cohort-aggregated ATT estimates on the vertical axis. A pattern compatible with the design consists of substantively small pre-treatment effects without a systematic trajectory, followed by post-treatment changes after SSRI can plausibly influence production. Monotonic improvement before treatment would raise concerns about anticipatory behavior, treatment targeting, or nonparallel trends. A discontinuity consistently appearing before the recorded date would suggest date misclassification.

The headline causal result should then report the cell-weighted overall ATT from the Callaway-Sant'Anna doubly robust estimator using not-yet-treated controls. Table 2 should contain the estimate, cluster-robust standard error, confidence interval, comparison group, event horizon, and sample support. A never-treated-only estimate should be reported where sufficient never-treated cells remain. A PO-volume-weighted ATT can be included for managerial interpretation but must be labeled as a different estimand because it gives greater influence to high-volume cells.

H1 predicts a negative ATT. The completed paper should report both the absolute percentage-point effect and the relative change from the untreated baseline, without conflating those scales.

Dynamic results should distinguish short-run and medium-run effects. A delayed reduction beginning around production activation would support the proposed requirements-to-workflow sequence. An initial deterioration followed by improvement could indicate implementation disruption and subsequent learning. A temporary improvement that quickly disappears could be more consistent with managerial attention than durable rule redesign. Growing effects could indicate cumulative rule correction and organizational learning. The data, rather than a preferred narrative, determine which interpretation applies.

Mechanism results should apply the same estimator to first-pass completion, requirements-defect escape, manual handoffs, repeated transitions, and post-release corrections. H2 predicts higher first-pass completion and fewer routing interventions; H3 predicts lower defect escape; H4 predicts fewer manual handoffs and repeated states. A temporally ordered pattern in which requirements-defect escape changes before downstream exception rates would be especially consistent with the proposed mechanism. Nevertheless, the manuscript should call these mechanism-consistent treatment effects rather than an identified mediation decomposition.

Heterogeneous treatment effects should be presented in a dedicated table. For each moderator, report low-, medium-, and high-baseline subgroup ATTs and the direct contrast between high and low groups. H6 predicts a larger absolute effect under high workflow complexity; H7 predicts a larger effect under greater supplier heterogeneity; H8 predicts stronger effects under greater system fragmentation until technical constraints become binding; and H9 predicts larger effects under higher demand volatility. The inferential focus is the subgroup difference itself. "Significant in one group but not another" is not valid evidence of heterogeneity.

Negative-control outcomes are critical because they test causal scope. The paper should contrast workflow routing failures, requirements-related incidents, first-pass completion, and manual rework with outcomes such as exogenous supplier production shutdowns and independent carrier delays. If SSRI appears to improve outcomes that it cannot plausibly influence, the analysis should investigate broader contemporaneous performance programs or selection effects.

Placebo interventions provide a second falsification layer. Pseudo-treatment dates should be assigned strictly before actual adoption while preserving the structure of the estimation window. The complete estimator should be rerun. Repeated large placebo effects would indicate that the method is capturing recurrent trend breaks rather than SSRI. The paper should report the placebo distribution and the position of the actual estimate relative to it.

Estimator robustness should compare the primary Callaway-Sant'Anna ATT with Sun-Abraham interaction-weighted estimates and Borusyak-Jaravel-Spies imputation estimates. A matched risk-set DiD can be added to improve transparency of the comparison

population. Conventional TWFE may appear only as a benchmark. Agreement across modern estimators strengthens numerical robustness but does not replace the identifying assumptions.

Comparison-group robustness should use not-yet-treated controls, never-treated controls where available, matched not-yet-treated controls, same-domain controls, and same-PO-type-family controls. Material variation across these estimates is informative about counterfactual sensitivity and should be reported rather than hidden. Alternative treatment timing should similarly re-index event time to mandate, first completed synthesis gate, and production activation. A credible operational effect should not depend on choosing the most favorable date after inspecting the outcome.

Co-intervention analysis is identification-critical. The clean-rollout specification should remove treatment cells that simultaneously undergo major ERP migration, supplier-platform conversion, and master-data cleansing, organizational restructuring, or unrelated automation. If the headline ATT disappears, the original estimate should not be attributed exclusively to SSRI. Stable-supplier and stable-PO-type samples further test whether compositional turnover rather than workflow redesign explains the result.

Balanced-horizon and leave-one-cohort-out analyses should verify that long-run results are not artifacts of changing cohort composition or one dominant adoption wave.

The final robustness layer is sensitivity to parallel-trends violations. Rambachan-Roth analysis should report how large a deviation from the extrapolated pre-treatment trend is necessary to overturn the substantive conclusion. A result that remains informative under meaningful deviations is more credible than one that requires exact parallel trends. The results section should therefore report the conventional interval, sensitivity restriction, robust interval, and breakdown value.

Strong causal language is permitted only when treatment timing is institutionally defensible, pre-treatment deviations are substantively limited, post-treatment dynamics are temporally plausible, preferred staggered estimators are consistent, negative controls remain approximately null, placebo dates do not reproduce the effect, clean-rollout results persist, no single cohort dominates, and conclusions survive reasonable deviations from exact parallel trends. If these conditions are only partially met, the correct conclusion is that the estimates are consistent with a causal effect under stated assumptions. If treatment timing clearly responds to unobserved outcome deterioration or major co-interventions are inseparable, the paper should downgrade its claim to an adjusted association rather than using causal language.

5. Discussion and Conclusion

The central theoretical contribution of this study is to treat requirements integration as a causal organizational intervention rather than merely an attribute associated with successful projects. The distinction matters because stakeholder presence and stakeholder synthesis are not equivalent. A retailer can involve merchandising, procurement, logistics, finance, supplier operations, and engineering while still allowing each function to define requirements independently. In that environment, every local specification can be defensible while the end-to-end workflow remains inconsistent. SSRI attempts to solve this problem by creating a governance point at which terminology, decision rules, exceptions, dependencies, PO-type differences, and acceptance criteria must be reconciled before they become executable workflow logic.

If future estimation yields a credible negative ATT on workflow-induced exceptions, the causal interpretation would be narrow but important: operating cells subjected to mandatory SSRI experienced fewer workflow failures than they would have experienced under the estimated untreated trajectory. The result would not establish that every stakeholder workshop is valuable, that all PO types respond equally, or that the same magnitude applies to every retailer. It would establish an intervention effect for treated cells under the stated identification assumptions.

A null effect would also be scientifically informative. Existing workflows may already be sufficiently mature; technical constraints may dominate requirements quality; implementation fidelity may be weak; or benefits may be concentrated only in complex workflows and disappear in an aggregate ATT. Similarly, a credible increase in exceptions would challenge simple collaboration narratives. Cross-functional governance might increase process complexity, tighten validations, expose previously hidden incidents, or cause temporary implementation disruption. Dynamic event studies are therefore essential for distinguishing durable harm from short-run transition costs.

The framework contributes to requirements engineering by emphasizing three constructs. First, stakeholder coverage is incomplete without reconciliation. Second, cross-type coherence matters when workflow types share data and rules. Third, defect escape provides a bridge between requirements quality and operational performance. The economically consequential defect is not merely a flawed requirement but a flaw that survives review and testing and reaches production transactions.

The study also provides a micro-foundation for supply-chain integration research. Integration depends not only on whether information is shared but on whether systems agree about which information is required, when it is authoritative, what conditions govern it, which exceptions are valid, and which stakeholder possesses decision rights. Requirements integration therefore precedes many operational phenomena later described as supply-chain coordination or information integration.

Managerially, organizations should organize PO requirements around end-to-end scenarios rather than functional documents alone. For each PO type, governance should explicitly map initiation, required data, authorization, release, acknowledgement, amendment, fulfillment, receiving, matching, cancellation, and exception behavior. The SSRI treatment definition also supplies an operational checklist: stakeholder coverage, PO-type decomposition, conflict resolution, traceability, scenario validation, and sign-off must be auditable rather than assumed.

The heterogeneity framework has direct resource-allocation implications. Intensive synthesis consumes stakeholder time and engineering capacity, so uniform implementation may be inefficient. If H6-H9 are supported, the highest-value targets are likely to be workflows with numerous branches, heterogeneous supplier capabilities, frequent amendments, many technical interfaces, and high demand volatility. Low-complexity routine flows may require lighter governance. This converts requirements engineering from a universal process prescription into a risk-adjusted design decision.

The design also clarifies what previous correlational findings can and cannot establish. Observational relationships between integration and performance are useful for theory development, construct validation, and mechanism discovery, but they cannot by themselves answer whether introducing an integration practice will improve performance. If the causal ATT is smaller than prior associations, omitted managerial capability or selection may explain part of the earlier relationship. If effects are highly heterogeneous, pooled correlations may hide where requirements integration is actually valuable.

Several limitations remain. First, treatment is not randomized. Conditional parallel trends concerns an unobservable untreated potential outcome and can never be proven. Documentary evidence on rollout timing, pre-treatment dynamics, falsification outcomes, and sensitivity bounds can strengthen but not eliminate this limitation. Second, SSRI may be implemented as a bundle containing facilitation, testing, training, documentation, and heightened managerial attention. If those elements cannot be separated, the ATT is an effect of the SSRI implementation package rather than an isolated effect of cross-functional discussion. Third, knowledge and workflow templates may spill across cells, contaminating controls. Fourth, better documentation may change measured incident rates even if underlying failure rates do not. Fifth, a treatment effect from one enterprise has limited transportability to retailers with different supplier networks, architectures, or governance maturity.

Future research should pursue stepped-wedge cluster randomization of SSRI rollout and component-level experiments that isolate stakeholder coverage, conflict resolution, traceability, and scenario testing. Multi-retailer replication could then test transportability across different architectures and supplier networks.

Economic evaluation is another important extension. The relevant decision is not merely whether SSRI reduces exceptions but whether the avoided cost exceeds the intervention cost. A complete model would combine avoided exception resolution, avoided delay, avoided rework, and prevented downstream mismatch costs, then subtract stakeholder time, facilitation, testing, and engineering costs. Causal outcome estimates can therefore become inputs into a transparent investment calculation rather than remaining abstract statistical results.

In conclusion, multi-type PO workflows are executable representations of interdependent commercial, logistical, financial, supplier, and technical decisions. When these decisions are elicited separately, local correctness can coexist with system-level failure. SSRI is proposed as a structured intervention for resolving those dependencies before requirements become production rules. The paper's central question is therefore counterfactual: what would workflow performance have been for treated operating cells had SSRI

not been introduced? A staggered difference-in-differences design, implemented with cohort-time treatment effects, modern event-study estimators, placebo treatments, negative controls, co-intervention restrictions, spillover analysis, and sensitivity to nonparallel trends, provides a rigorous strategy for answering that question. No numerical causal claim is made without the required retailer microdata. The scientific contribution lies in specifying a design under which a future numerical claim would be credible.

References

- [1] Abadie, A. (2005). Semiparametric difference-in-differences estimators. *The Review of Economic Studies*, 72(1), 1–19. <https://doi.org/10.1111/0034-6527.00321>
- [2] Bano, M., & Zowghi, D. (2015). A systematic review on the relationship between user involvement and system success. *Information and Software Technology*, 58, 148–169. <https://doi.org/10.1016/j.infsof.2014.06.011>
- [3] Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics*, 119(1), 249–275. <https://doi.org/10.1162/003355304772839588>
- [4] Borusyak, K., Jaravel, X., & Spiess, J. (2024). Revisiting event-study designs: Robust and efficient estimation. *The Review of Economic Studies*, 91(6), 3253–3285. <https://doi.org/10.1093/restud/rdae007>
- [5] Callaway, B., & Sant'Anna, P. H. C. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2), 200–230. <https://doi.org/10.1016/j.jeconom.2020.12.001>
- [6] Daft, R. L., & Lengel, R. H. (1986). Organizational information requirements, media richness and structural design. *Management Science*, 32(5), 554–571. <https://doi.org/10.1287/mnsc.32.5.554>
- [7] Damian, D., & Chisan, J. (2006). An empirical study of the complex relationships between requirements engineering processes and other processes that lead to payoffs in productivity, quality, and risk management. *IEEE Transactions on Software Engineering*, 32(7), 433–453. <https://doi.org/10.1109/TSE.2006.61>
- [8] Davila, A., Gupta, M., & Palmer, R. (2003). Moving procurement systems to the internet: The adoption and use of e-procurement technology models. *European Management Journal*, 21(1), 11–23. [https://doi.org/10.1016/S0263-2373\(02\)00155-X](https://doi.org/10.1016/S0263-2373(02)00155-X)
- [9] de Chaisemartin, C., & D'Haultfoeulle, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9), 2964–2996. <https://doi.org/10.1257/aer.20181169>
- [10] Dyer, J. H., & Singh, H. (1998). The relational view: Cooperative strategy and sources of interorganizational competitive advantage. *Academy of Management Review*, 23(4), 660–679. <https://doi.org/10.5465/amr.1998.1255632>
- [11] Flynn, B. B., Huo, B., & Zhao, X. (2010). The impact of supply chain integration on performance: A contingency and configuration approach. *Journal of Operations Management*, 28(1), 58–71. <https://doi.org/10.1016/j.jom.2009.06.001>
- [12] Galbraith, J. R. (1974). Organization design: An information processing view. *Interfaces*, 4(3), 28–36. <https://doi.org/10.1287/inte.4.3.28>
- [13] Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2), 254–277. <https://doi.org/10.1016/j.jeconom.2021.03.014>
- [14] Hofmann, H. F., & Lehner, F. (2001). Requirements engineering as a success factor in software projects. *IEEE Software*, 18(4), 58–66. <https://doi.org/10.1109/MS.2001.936219>
- [15] International Organization for Standardization, International Electrotechnical Commission, & Institute of Electrical and Electronics Engineers. (2018). ISO/IEC/IEEE 29148:2018 Systems and software engineering—Life cycle processes—Requirements engineering. ISO.
- [16] Kotonya, G., & Sommerville, I. (1998). Requirements engineering: Processes and techniques. John Wiley & Sons.
- [17] Lim, S. L., Quercia, D., & Finkelstein, A. (2010). StakeNet: Using social networks to analyse the stakeholders of large-scale software projects. In *Proceedings of the 32nd ACM/IEEE International Conference on Software Engineering* (pp. 295–304). ACM. <https://doi.org/10.1145/1806799.1806844>
- [18] Méndez Fernández, D., Wagner, S., Kalinowski, M., Felderer, M., Mafrá, P., Vetrò, A., Conte, T., Christiansson, M.-T., Greer, D., Lassenius, C., Männistö, T., Nayeib, M., Oivo, M., Penzenstadler, B., Pfahl, D., Prikładnicki, R., Ruhe, G., Schekelmann, A., Sen, S., & Wieringa, R. (2017). Naming the pain in requirements engineering: Contemporary problems, causes, and effects in practice. *Empirical Software Engineering*, 22(5), 2298–2338. <https://doi.org/10.1007/s10664-016-9451-7>
- [19] Prajogo, D., & Olhager, J. (2012). Supply chain integration and performance: The effects of long-term relationships, information technology and sharing, and logistics integration. *International Journal of Production Economics*, 135(1), 514–522. <https://doi.org/10.1016/j.ijpe.2011.09.001>
- [20] Rambachan, A., & Roth, J. (2023). A more credible approach to parallel trends. *The Review of Economic Studies*, 90(5), 2555–2591. <https://doi.org/10.1093/restud/rdado18>
- [21] Roth, J. (2022). Pretest with caution: Event-study estimates after testing for parallel trends. *American Economic Review: Insights*, 4(3), 305–322. <https://doi.org/10.1257/aeri.20210236>
- [22] Roth, J., Sant'Anna, P. H. C., Bilinski, A., & Poe, J. (2023). What's trending in difference-in-differences? A synthesis of the recent econometrics literature. *Journal of Econometrics*, 235(2), 2218–2244. <https://doi.org/10.1016/j.jeconom.2023.03.008>
- [23] Sant'Anna, P. H. C., & Zhao, J. (2020). Doubly robust difference-in-differences estimators. *Journal of Econometrics*, 219(1), 101–122. <https://doi.org/10.1016/j.jeconom.2020.06.003>
- [24] S. K. Gunda, "Analyzing Machine Learning Techniques for Software Defect Prediction: A Comprehensive Performance Comparison," 2024 Asian Conference on Intelligent Technologies (ACOIT), KOLAR, India, 2024, pp. 1–5, <https://doi.org/10.1109/ACOIT62457.2024.10939610>.

- [25] Sharp, H., Finkelstein, A., & Galal, G. (1999). Stakeholder identification in the requirements engineering process. In Proceedings of the 10th International Workshop on Database and Expert Systems Applications (pp. 387–391). IEEE. <https://doi.org/10.1109/DEXA.1999.795198>
- [26] Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2), 175–199. <https://doi.org/10.1016/j.jeconom.2020.09.006>
- [27] Tai, Y.-M., Ho, C.-F., & Wu, W.-H. (2010). the performance impact of implementing web-based e-procurement systems. *International Journal of Production Research*, 48(18), 5397–5414. <https://doi.org/10.1080/00207540903117915>
- [28] Tushman, M. L., & Nadler, D. A. (1978). Information processing as an integrating concept in organizational design. *Academy of Management Review*, 3(3), 613–624. <https://doi.org/10.5465/amr.1978.4305791>
- [29] Vaidyanathan, G., & Devaraj, S. (2008). The role of quality in e-procurement performance: An empirical analysis. *Journal of Operations Management*, 26(3), 407–425. <https://doi.org/10.1016/j.jom.2007.08.004>
- [30] Vickery, S. K., Jayaram, J., Droge, C., & Calantone, R. (2003). The effects of an integrative supply chain strategy on customer service and financial performance: An analysis of direct versus indirect relationships. *Journal of Operations Management*, 21(5), 523–539. [https://doi.org/10.1016/S0272-6963\(03\)00061-](https://doi.org/10.1016/S0272-6963(03)00061-)