

Conversion Rate Optimization Process: Stages, Roles, and Deliverables

- CRO as a disciplined experimentation system, not one-off tactics
- Evidence compounds: each test is an investment in learning
- Five-stage spine: Audit, Research, Hypothesize, Test, Analyze
- Every stage has an owner, an output, and an exit criterion
- Core terms: primary metric, guardrail, MDE, SRM, hypothesis record
- Artifact trail: funnel baseline, backlog, pre-registration, QA log, results memo



BETTER
EXPERIENCES
THROUGH
EVIDENCE

Why the Process Beats the Tactic

FROM NOISE
TO PROGRESS



-  Jumping straight to testing yields random results — the failure is process, not testing
-  Sequence compounds gains: audit, hypothesize, pre-register, launch, then bank learning
-  CRO is a loop: analysis from one experiment seeds the next round of hypotheses
-  “Purchase” and “lead” must mean the same thing across marketing, analytics, and CRM
-  Maturity stages: ad hoc, reactive, structured, strategic, embedded — most programs sit at structured
-  Madence keeps the loop alive: weekly experiment review, monthly roadmap, quarterly reset

Stage 1: Audit Measurement and Provenance Before Optimizing



Validate events before optimizing:

fire manually, confirm once, right value, right state



Check funnel integrity:

modals, SPA route changes, in-page micro-conversions



Client-side and server-side validation are separate but connected workflows



2026 stack: pixel for speed, server-side for durability, warehouse as source of record



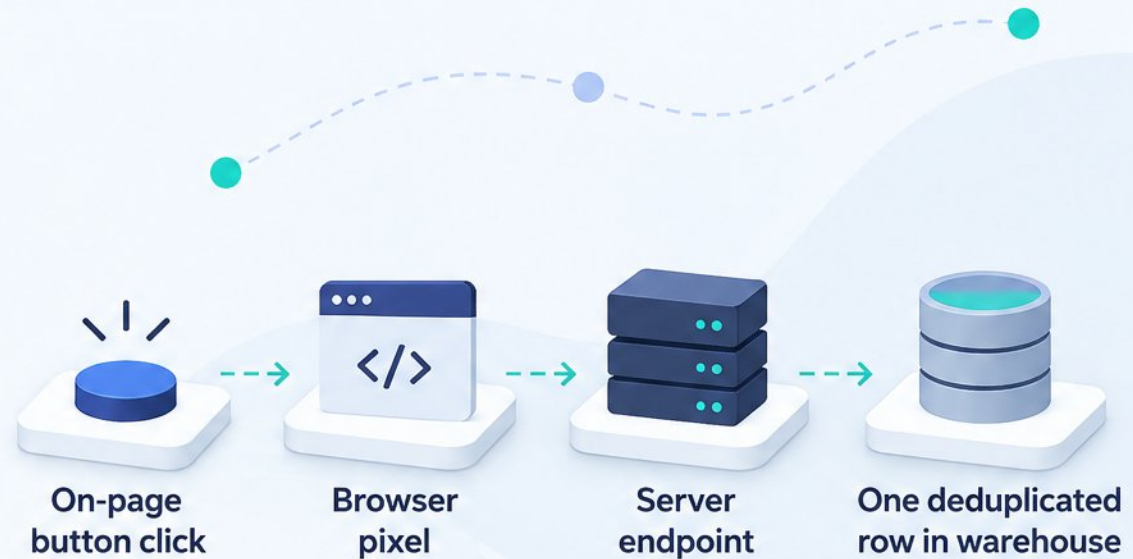
Hygiene: UTM conventions, click ID persistence, cross-domain, consent on every event



Failures: missing fields, double firing, post-assignment bot filtering, consent drops



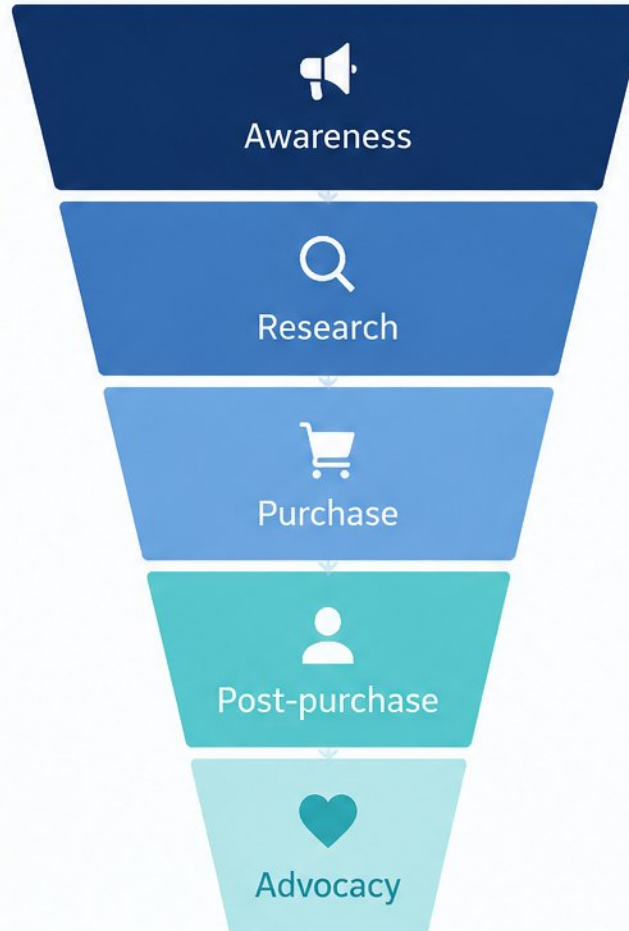
Deliverable: tracking plan plus measurement health verdict



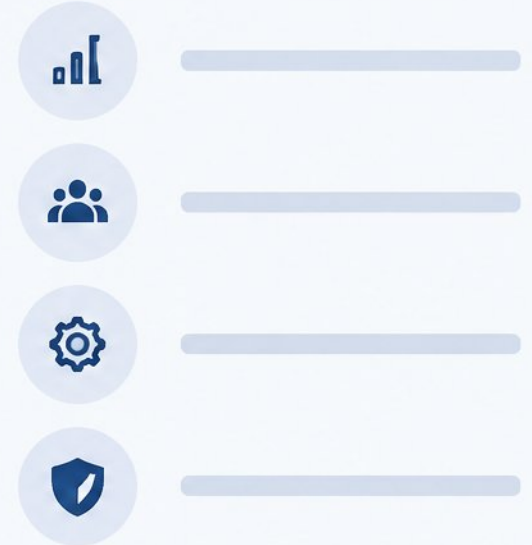
Mapping the Funnel and Sizing the Opportunity

- Map awareness to purchase and post-purchase; attach one primary metric per stage (W1).
- Pick one primary metric per stage plus three to five guardrail metrics (W1, W3).
- Keep funnels to five to eight steps; more steps dilute data and shrink power (W1).
- Connect drop-off to money: show revenue lost per step, not just headcount left (W1).
- Segment before concluding: mobile vs desktop, new vs returning, channel-specific paths (W1).
- Size opportunities by traffic volume, drop-off size, revenue exposed, and strategic fit.
- Deliverable: funnel map with stage rates, biggest leak, and ranked first-test shortlist.

From Awareness to Advocacy



Guardrail Metrics



The Funnel Spine



Stage 2: Research and Insight, Quantitative and Qualitative

Quantitative Evidence

Tells you what and where



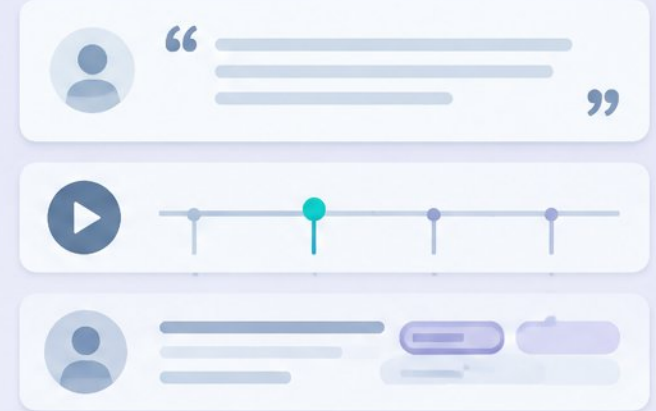
Funnel drop-off, analytics segmentation, heatmaps, form analytics



Shared
Friction Point

Qualitative Evidence

Tells you why



User testing, session recordings, surveys, customer interviews, chat logs

- **Research** is a systematic search for user behavior; it produces hypotheses, not opinions
- **Quantitative methods** tell you what and where: funnel drop-off, analytics segmentation, heatmaps, form analytics
- **Qualitative methods** tell you why: user testing, session recordings, surveys, customer interviews, chat logs
- **Triangulate:** a heuristic finding stays an "area of interest" until confirmed by another method
- **Cross-reference both evidence types:** a big drop-off plus a matching qualitative signal beats a hunch
- **Deliverable:** insight repository plus issue classification — instrument, just do it, test, hypothesize, investigate

Artifact Trail



Stage 3: Writing Hypotheses That Can Be Falsified

- “Because we observed [evidence], we believe [change] will cause [outcome] for [audience]”
- Four required parts: observation, single change, measurable outcome, success threshold
- The mechanism, not the prediction, makes wins reusable and losses instructive
- Isolate one meaningful change per variant — bundled changes are uninterpretable
- Kill vanity hypotheses: no mechanism, no evidence, and 500% lift wishes
- Deliverable: a hypothesis record with evidence links, expected lift, and mechanism

Hypothesis Record



Observation



Single Change



Measurable Outcome



Decision Threshold



Evidence Link



Tracking Plan



Insight Repository



Hypothesis Record



Pre-registration Brief



QA Log



Results Memo

Pre-Registration: Sample Size, Duration, and Decision Rules

1

Lock the plan before launch:
hypothesis, metrics, sample size,
duration, decision rule.

2

Sample size needs four inputs:
baseline rate, MDE, alpha 0.05,
power 0.80.

3

Halving the effect to detect
roughly quadruples the sample
needed.

4

Run to sample size plus one full
cycle: 7 days minimum, 4–6 weeks
maximum.

5

Define guardrail violation
thresholds up front, beside the
ship/iterate/kill rule.



Pre-Registration Plan



Hypothesis



Metrics



Sample size



Duration



Decision rule



Sample Size Model



Baseline rate



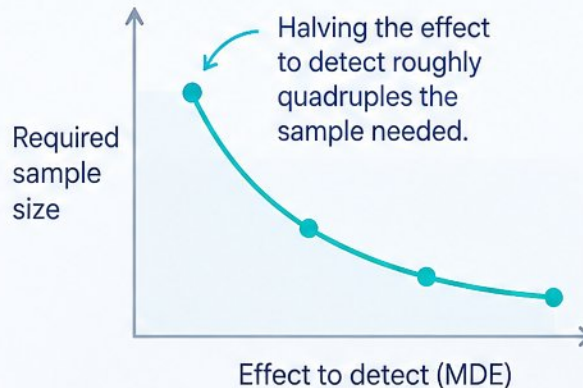
MDE



Alpha 0.05



Power 0.80



Decision rule

Ship

Iterate

Kill



Guardrail violation
thresholds up front.

Prioritization Frameworks: ICE, PIE, PXL, and Hybrids

Different approaches. A common goal: focus on what creates the most value.

ICE

Impact, Confidence, Ease



Fastest, drifts with enthusiasm unless calibrated

PIE

Potential, Importance, Ease



Built for CRO, forces headroom and page value

PXL

Binary, evidence-anchored



Binary, evidence-anchored questions reduce HiPPO bias but need research

Hybrid

ICE to triage, PXL to pressure-test



Audit



Research



Hypothesize



Test



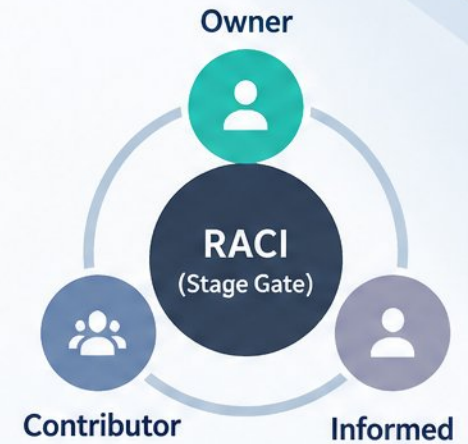
Analyze

- All frameworks weigh upside, evidence, and effort, differing in subjectivity
- ICE: Impact, Confidence, Ease — fastest, drifts with enthusiasm unless calibrated
- PIE: Potential, Importance, Ease — built for CRO, forces headroom and page value
- PXL: binary, evidence-anchored questions reduce HiPPO bias but need research
- Hybrid: ICE to triage the backlog, PXL to pressure-test the shortlist
- After ranking, check power and feasibility; higher-traffic ideas jump the queue
- Re-score after every test: winners raise confidence, losers lower it

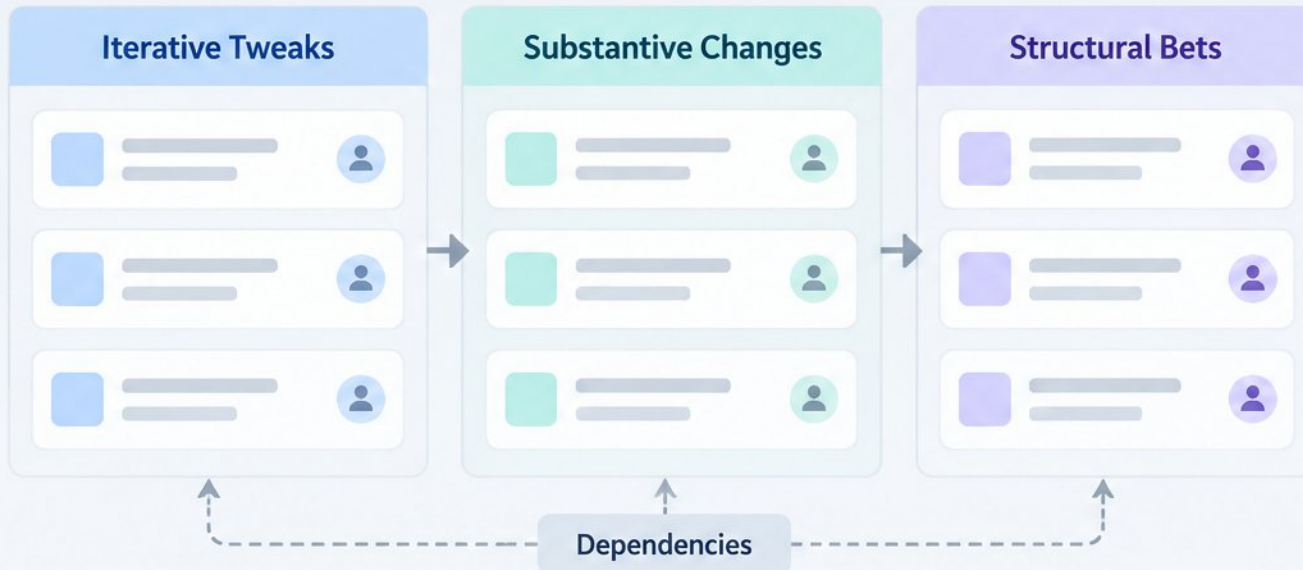
BETTER IDEAS
BRIGHTER OUTCOMES

Building and Governing the Experiment Roadmap

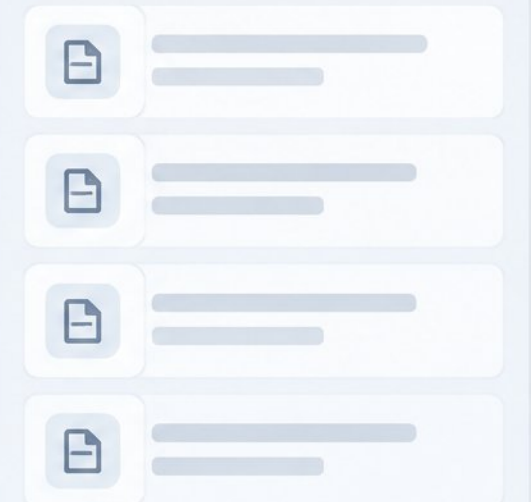
- Score every hypothesis on consistent criteria; name evidence sources, score independently (W2)
- Balance iterative tweaks, substantive changes, and larger structural bets (W10, W2)
- Velocity: 2–4 structured tests/month for lean teams; 8–12 concurrent for high-traffic sites (W3)
- Quick wins fund credibility; structural tests and platform work carry compounding value (W3)
- Register every active test: name, pages affected, traffic allocation—check conflicts pre-launch (W10)
- Deliverable: rolling backlog and quarterly roadmap with scores, owners, and dependencies (W2)



Quarterly Experiment Roadmap (Rolling)



Conflict Check Log



Stage 4: Execution, Build Paths, and Launch QA

Client-side testing



- Client-side tools deploy fast on marketing pages
- Flicker risk
- Added JS weight

Server-side and feature flags



- Server-side and feature flags remove flicker, handle backend logic, more reliable
- Engineering dependency



QA Checklist

- ✓ Run an A/A test on new surfaces; SRM means broken instrumentation
- ✓ Log exposure when the user sees the variant, not on page load
- ✓ QA checklist: rendering, tracking, assignment stickiness, bots, rollback plan



Deliverable: completed QA log with validation date and launch sign-off

Artifact trail



Tracking plan



Insight repository



Hypothesis record



Pre-registration brief



QA log



Results memo

Sample Ratio Mismatch and Live-Test Integrity

TRUSTED EXPERIMENTS. RELIABLE DECISIONS.



SRM: intended vs. observed split deviates significantly — a validity gate, not an effect



Chi-squared test on users, not sessions;
alarm threshold $p < 0.001$



Check within 48 hours of launch,
then monitor daily



Common causes: hashing bugs, client-side
exposure loss, redirects, caching,
post-assignment bot filtering



Segment by browser, device, geo, day;
compare split at assignment, exposure,
analysis

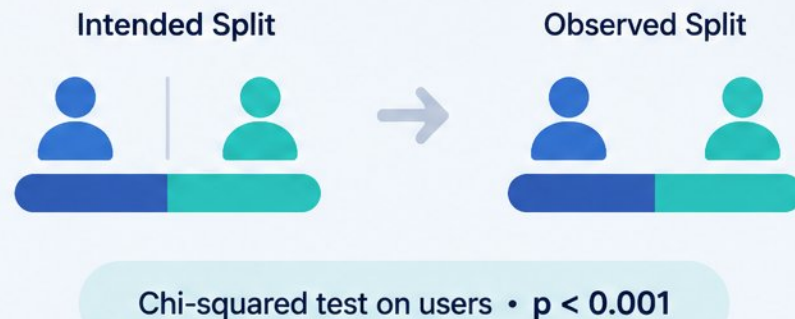


Fix the filter to recover lost data;
re-randomize and restart if it was never captured

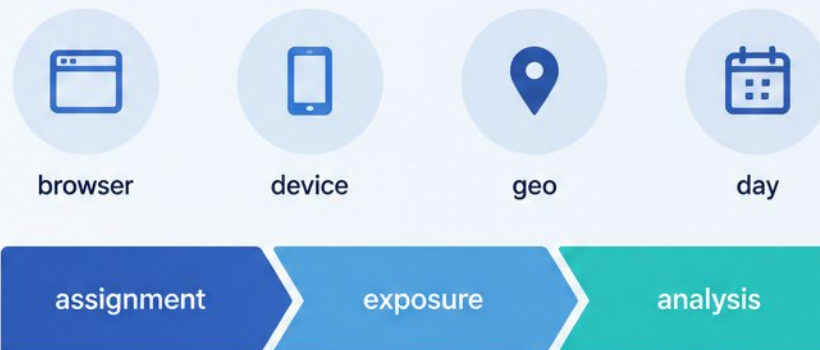


SRM fires: the lift is not a weaker result —
it is not a result

Split Integrity Check



Investigate and Segment



Stage 5: Reading Results Without Fooling Yourself



Quality gates first: SRM check ($p < 0.001$), sample size reached, then primary metric



Match method to workflow: fixed-horizon frequentist, Bayesian, or sequential



Flat-prior Bayes does not fix peeking; it reproduces naive frequentist error rates



Report effect estimate, interval, sample sizes, and guardrails together



Segment wins for device, source, new vs. returning; watch novelty effects



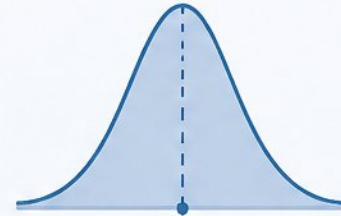
Decide: ship, iterate, stop, or re-test. Document losses as evidence

Choose the right analytical lens

Fixed-Horizon Frequentist

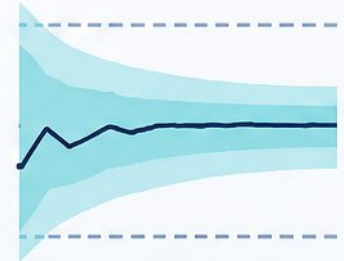


Bayesian



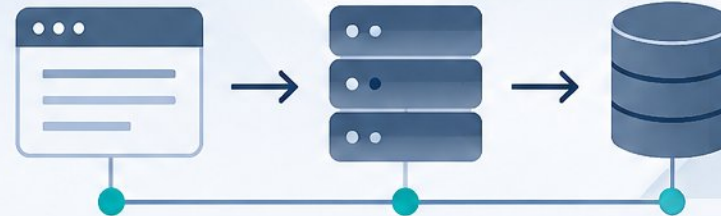
Flat-prior Bayes does not fix peeking; it reproduces naive frequentist error rates

Sequential



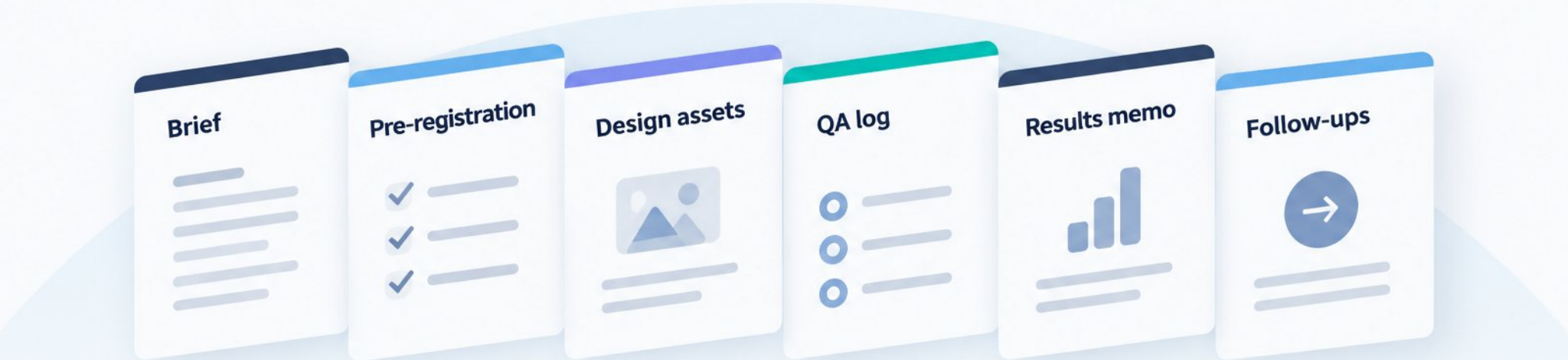
Always-valid monitoring-proof intervals

From data to a confident decision



A clear signal path from event to insight





- Standard deliverable set: brief, pre-registration, design assets, QA log, results memo, follow-ups
- Results record: hypothesis, control and variant rates, relative lift, confidence, decision
- One master test log: one row per experiment, with implemented and revenue impact columns
- Compound deployed relative lifts; sequential gains multiply, they do not add
- Report qualified and all-deployed compounded uplift side by side
- Document null and losing results as thoroughly as wins

[illegible]

Roles, RACI, Operating Cadence, and the 90-Day Start

ALIGN PEOPLE
IMPROVE HANDOFFS
DRIVE IMPACT

RACI Across the Five Handoff Gates

Clear ownership. Smoother handoffs. Stronger outcomes.

Role	Discover	Define	Define	Build	Launch
 CRO lead	A	C	I	I	C
 growth PM	C	A	C	C	I
 researcher	R	C	I	I	I
 designer	I	C	R	C	I
 developer	I	I	C	R	C
 analyst	I	I	C	C	R

A

 Accountable

R

 Responsible

C

 Contributor

I

 Informed

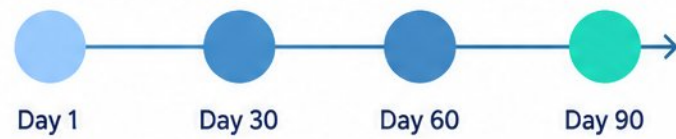
Operating Cadence

A steady rhythm from experiments to long-term progress.



The 90-Day Start

Turn alignment into action.



Key Takeaways

-  Map core roles: CRO lead, growth PM, researcher, designer, developer, analyst
-  Every stage needs one accountable owner and five explicit handoff gates
-  Cadence: weekly 30-minute experiment review, monthly roadmap, quarterly reset
-  Friction is structural, not personal: fix handoffs and ownership, not people
-  Score maturity across goals, process, culture, and data — fix the constraint
-  Knowledge base is the strongest predictor of maturity; commit to one lift
-  Each role leaves with a Monday-ready action and a shared 90-day plan

PersonWise.

PersonWise • AI digital-human course platform



Scan the code or click anywhere to watch this course live, with voice Q&A

personwise.ai/t/82cecac2/public