GROWTHLOOP

Experiment-Level Randomization vs. Global Control

Exec summary · Experimentation design · July 2026

GrowthLoop’s experiment-level control group

GrowthLoop implementation: each journey/audience runs its own audience-level RCT, as opposed to a global control group.

How randomization works within GrowthLoop

- Assignment uses a consistent (deterministic) hash of the user ID plus an audience-specific salt
- It builds a stable snap_hash from the audience’s unique keys, e.g. user id or customer id
- It assigns control when positive_hash % 100 < control_percentage * 100

This means, assignment experimentation is

- Deterministic and stateless: the same user always lands in the same bucket for a given audience
- Decorrelated across campaigns: a user's bucket for audience A is statistically independent of their bucket for audience B

GrowthLoop's audience-level design produces an unbiased estimate of campaign lift, conditional on the SUTVA assumption covered below. Because assignment is a pure hash of the user ID and a campaign-specific salt, test and control users are probabilistically equivalent by construction — there's no inclusion criterion tied to user characteristics or behavior that could correlate with the outcome and introduce bias.

This design allows for causal measurement at any given point in time over the marginal campaign in question, and does not suffer from temporal misalignment, which can arise when a control group taken prior to marketing interventions is measured against a subsequent intervention. Here, every audience or journey draws its own fresh, independent split at the moment it activates, so the resulting estimate is causal at any point in the program's life and absorbs cumulative effects from prior campaigns without needing separate modeling.

Statistical power requirements are unchanged from any standard test/control design — GrowthLoop's approach needs the same sample sizes, in expectation, as any properly randomized experiment. The

trade-off with global hold-outgroups is in the distribution of the effect: experiment-level splits will produce the same aggregate ROI across a larger number of partially affected participants,.

The threat of SUTVA violation in GrowthLoop's experiment-level control group

SUTVA describes a threat to causal inference in which treatment "spills" over to members that were assigned to the control group. It is helpful to differentiate two sources of SUTVA violation:

(1) **Within-experiment network interference** — one user's treatment leaks to another user *in the same experiment* through a direct connection (a friend request, a share button, a referral);
(2) **Cross-experiment interaction** — does *Campaign/Test A* contaminate the result of a completely different, concurrently-running *Campaign/Test B* when both draw independently from the same population?

Note that the second variant of SUTVA violation is infinitely more likely – users get assigned to the control group for one export to Meta, but a subset of these users gets assigned to the *treatment* group for a different, similar export conducted at the same time.[1] This scenario becomes a causal threat if and only if assignment to the treatment group of the concurrent experiment has an *interactive* effect on the outcome of the initial experiment.

Example 1: A subset of users is assigned to the control group for a Meta ad campaign increasing their AOV, while simultaneously being assigned to the treatment group of an email experiment maximizing sign-up for family members through the audience-level hash. Most likely, the treatment assignment in the secondary campaign will be orthogonal to the outcome of the primary campaign.

Example 2: A subset of users is assigned to the control group for a Meta ad campaign increasing their AOV, while simultaneously being assigned to the treatment group of a Google Ads campaign likewise aimed at increasing their AOV. Here, SUTVA is violated only insofar as the contagion effect is of different magnitude for the control and treatment group.

$$\text{Bias}_{FB} = \Pr(G_i{=}1)\Big[\big(Y_{i,\,FB=1,\,G=1} - Y_{i,\,FB=1,\,G=0}\big) - \big(Y_{i,\,FB=0,\,G=1} - Y_{i,\,FB=0,\,G=0}\big)\Big]$$

Bias for users independently hashed into the Google-treatment arm; FB = Meta; G=Google. If the AOV lift from Google exposure among users in the Facebook-treatment arm equals the AOV lift from Google exposure among users in the Facebook-control arm, the bias in the Facebook campaign's estimated effect is exactly zeroed.

In practice, negative examples where these deltas are *not homogeneous* are virtually non-existent. For

[1] https://blog.statsig.com/embracing-overlapping-a-b-tests-and-the-danger-of-isolating-experiments-cb0a69e09d3

example, researchers at Microsoft find no difference in the deltas across a large number of online experiments when results of the primary experiments are broken down by assignment to the secondary experiments.[2] The same findings have emerged from applied experiment work at Meta, and many other avantgard internet consumer companies.[3]

Global-level control groups

Global control relies on one shared, standing panel of users held out of every campaign, rather than a fresh split drawn for each one. Because there's no per-campaign randomization step, isolating any single campaign's effect from that shared panel either (1) requires constructing a matched or synthetic control after the fact — selecting comparison users based on similarity to the treated audience – a significant per-campaign effort; (2) or assumes that inclusion criteria into the audience or journey are not related to the outcome, which in practice is often violated.

Example: If a conversion rate optimization campaign is run against an audience subset of young customers with a higher baseline proclivity to convert, but the control group is built globally a priori, measurement is not causal.

The design is also vulnerable to drift over time. Because the panel is fixed and persistent rather than freshly drawn per campaign, if cumulative effects from earlier campaigns have already shifted the panel's (read: global control group) baseline behavior by the time a new campaign runs, the comparison is no longer against a clean, unaffected baseline — the resulting ROI figure stops reliably measuring a causal effect. Rotating panel designs – essentially refreshing the composition of the control group over time – are complex and hard to execute.

Example: If an email campaign aimed at maximizing AOV of a second order comes after a newly scaled first-order-recommendation flow, measurement against a global control group taken prior to that scaled flow will not be causal

Compared to experiment-level holdout groups, global holdout designs will produce the same aggregate ROI across a smaller number of fully affected participants across campaigns.

On interference, global control does close off cross-experiment contamination, since holdout members are excluded from every campaign rather than just one. It remains exposed to within-experiment network

---

[2] https://www.microsoft.com/en-us/research/articles/a-b-interactions-a-call-to-relax/

[3] https://blog.statsig.com/embracing-overlapping-a-b-tests-and-the-danger-of-isolating-experiments-cb0a69e09d3 For a summary

interference, though — a holdout member's spouse, friend, or household member being treated by some campaign can still influence that member's behavior, since this design, like audience-level randomization, doesn't account for social-network structure.

Summary: Global vs. experiment-level control group

	Audience-level (GrowthLoop)	Global control
Control group	Fresh per-audience/journey split	One shared standing panel
Assignment method	Consistent hash(user, audience/journey salt)	Matched/synthetic control per campaign
Bias	Unbiased under SUTVA assumption (below)	Biased if inclusion criteria of audience/journeys are related to the outcome
Durability	Can take into account cumulative effects and is causal at any point in time	Does not measure causal ROI if some cumulative effects of previous campaigns have moved the level of the control group over time
Power	Same power needed in expectation; possibly decay in overall ROI because of zero ROI for global control group	Same power needed; partial ROI for all program units
SUTVA risk	Both within-experiment network interference and cross-experiment interaction threats. The latter is hypothetical only (see detailed section above)	Within-experiment network interference risk

Recommendation: keep audience-level randomization as the default.

Sources: Gordon, Zettelmeyer, Bhargava & Chapsky (2019), *A Comparison of Approaches to Advertising Measurement: Evidence from Big Field Experiments at Facebook*, Marketing Science 38(2). Working paper PDF · Published version
Violations of SUTVA – Harvard IQSS Social Science Statistics Blog; Synthetic Controls Aren't Enough – Measured