

What a Retention Diagnostic hands you

This is a sample. Every account, number and chart in it is invented by a seeded Python script so you can see the shape and depth of the deliverables before you buy. It is not a client result, not anonymized client data, and not a benchmark. Re-running the script with the same seed (20261020) gives the same pack.

The made-up company

A subscription SaaS product with three plans (Starter and Team billed monthly, Business billed annually). 8,000 paying signups between Oct 2024 and Oct 2026; 5,439 paying accounts and \$18,203,124 ARR on the data cut (01 Oct 2026). It has account-level billing history and a weekly count of one core product action. That is all the Diagnostic needs to start.

Contents

Page	Deliverable	What it answers
2	D1 Churn definition memo	Which churn number are we arguing about?
3	D2 Survival by segment	When do accounts leave, and which ones?
4	D3 At-risk segments	Which groups are worth acting on, written as rules anyone can check?
5	D4 Comparison group test spec	What do we try, on whom, and how will we know?
6	D5 Executive brief (excerpt)	The one-page version for the person who signs off.
7	D6 Audience export	The list your team works from, plus the SQL behind it.
8	D7 Lift readout	Did it work? With an interval, and a verdict.
9	Power	How big a test has to be before it can tell you anything.
10	Methods and reproducibility	How this sample was made, and its limits.

How to read it

- In a real engagement the data is yours and the numbers are findings about your business. Here they are illustrations: the generator was built with the patterns the analysis then finds (see page 10), so nothing here shows that the method "works".
- The D7 outcome on page 8 is simulated, because a real readout needs a real test.
- The page count is trimmed. A real D5 runs 6–8 pages and D4 runs about 3.

D1 · CHURN DEFINITION MEMO

Three churn numbers, one standard

Accounts paying on 01 Oct 2025: 3,173 (\$10,383,336 ARR). Trailing-12-month window: 01 Oct 2025 to 01 Oct 2026.

Definition	Trailing 12 months	Rate
A. Logo churn, counted on cancellation-request date	829 accounts	26.1%
B. Logo churn, counted at end of paid term (standard)	876 accounts	27.6%
C. Gross revenue churn, ARR lost at end of term	\$2,208,228	21.3%

Recommendation: standardize on B

- Count a logo as churned when its paid term ends, not when it clicks cancel. Revenue stops at the end of term, and that is when a save is no longer possible.
- A and B differ by 47 accounts over the year here. The gap is timing: annual customers ask to cancel up to 45 days before renewal, and monthly customers finish the month they paid for. Reports that mix A and B will disagree by about that much every quarter.
- Report C next to B. In this sample revenue churn (21.3%) is lower than logo churn (27.6%) because the cheapest plan churns most. Quoting either alone gives a misleading picture.
- As a monthly-equivalent rate, B is 2.66% per month (compounded from the 12-month figure).

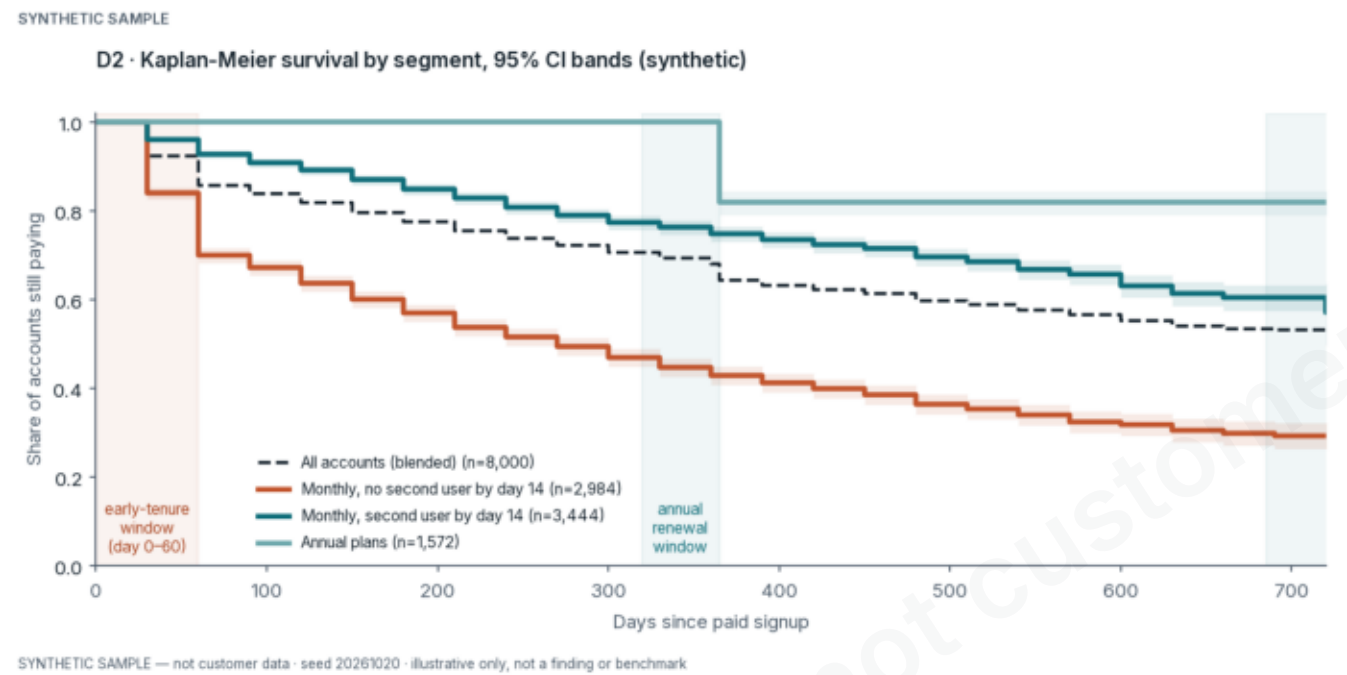
What gets signed off

- The definition (B) and its SQL, the 12-month window, and that reactivations within 30 days count as the same account (a choice; none occur in this sample).
- Nothing downstream (D2–D7) starts until this page is agreed.

```
-- D1 · definition B, trailing 12 months
with start_base as (
  select account_id, arr from accounts
  where signup_date <= :start and (ended_at is null or ended_at > :start))
select count(*) as accounts_at_start,
       count(*) filter (where a.ended_at > :start and a.ended_at <= :end) as churned_b,
       sum(s.arr) filter (where a.ended_at > :start and a.ended_at <= :end) / sum(s.arr) as revenue_churn_c
from start_base s join accounts a using (account_id);
```

D2 · SEGMENTED SURVIVAL ANALYSIS

Most of the loss happens early, or at renewal



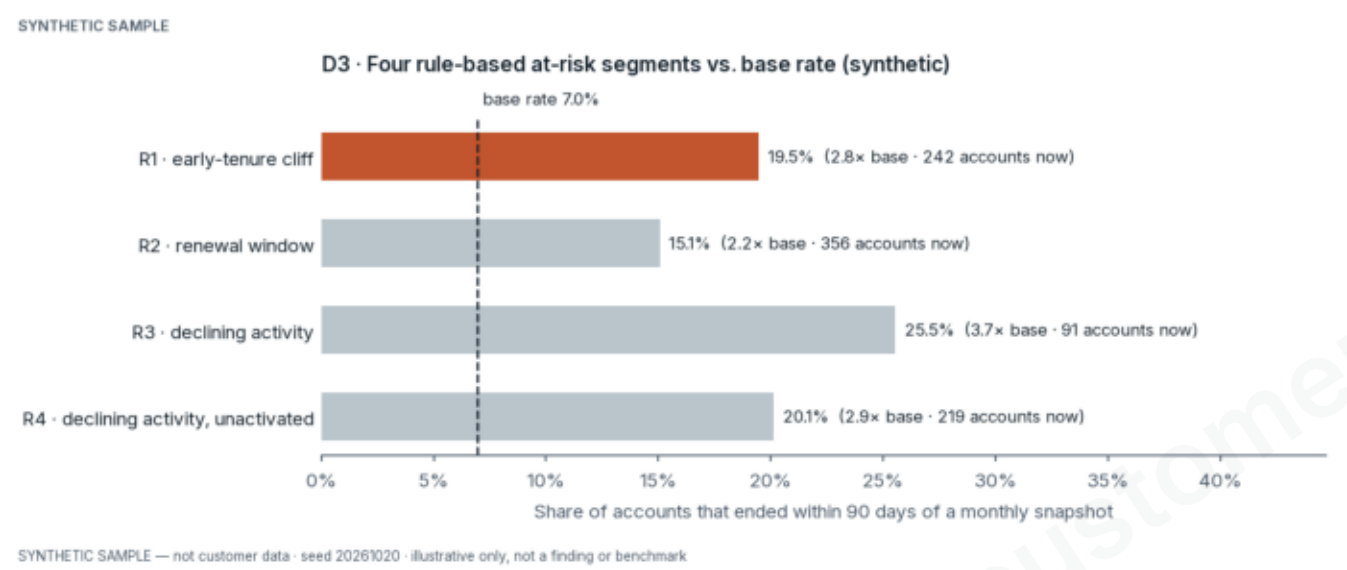
Segment	Accounts	Ended	Still paying at day 90	At day 365	Median life	Worst 30-day stretch
Monthly, no second user by day 14	2,984	1,531	67.2%	42.8%	270 days	days 30–60: 16.0% ended
Monthly, second user by day 14	3,444	805	90.8%	74.8%	not reached	days 0–30: 3.8% ended
Annual plans	1,572	145	100.0%	81.9%	not reached	days 360–390: 17.9% ended

What this would mean (if it were real data)

- Monthly accounts that never invite a second user in their first 14 days keep paying at much lower rates: 67% are still paying at day 90, against 91% for monthly accounts that do invite one. The gap opens fast in the first two months and keeps widening, more slowly, after that.
- Annual accounts barely move until renewal; the worst stretch is days 360–390. Mid-term saves are not where the money is for them.
- The blended curve (dashed) hides both patterns. Segment differences: log-rank test across the three segments, $p < 0.001$ (8,000 accounts, 2,481 ended).

D3 · EXPLAINABLE AT-RISK SEGMENTS

Four segments, each a rule you can check



	Rule (plain language)	Accounts now	90-day churn vs. base	ARR at risk, next 90 days	Confidence note
R1	Monthly plan, signed up self-serve, no second user invited in the first 14 days, under 90 days old	242	19.5% (2.8×)	\$44,887	4,791 account-snapshots; 23% of all churn. Earlier half 19%, later half 20%: holds in both.
R2	Annual plan, renewal date in the next 90 days	356	15.1% (2.2×)	\$487,942	2,085 account-snapshots; 8% of all churn. Earlier half 20%, later half 14%: holds in both.
R3	Any plan, 90+ days old, weekly core actions in the last 4 weeks down 40% or more vs. the 4 weeks before	91	25.5% (3.7×)	\$79,798	916 account-snapshots; 6% of all churn. Earlier half 27%, later half 25%: holds in both.
R4	Monthly plan, 90+ days old, no second user invited in the first 14 days, weekly core actions down 20% or more	219	20.1% (2.9×)	\$46,214	1,992 account-snapshots; 10% of all churn. Earlier half 23%, later half 19%: holds in both.

Base rate: 7.0% of accounts paying at a monthly snapshot ended within 90 days (19 snapshots, Jan 2025–Jul 2026, 57,711 account-snapshots). Support rule: at least 50 accounts today and at least 5% of churn, and the lift holds in both halves of the lookback. Segments can overlap (an account can match both R3 and R4); ARR at risk = today's ARR in the segment × its historical 90-day rate.

```
-- D3 · R3 declining activity
with w as (
  select account_id,
    avg(case when week_start >= current_date - 28 then core_actions end) as recent,
    avg(case when week_start < current_date - 28 then core_actions end) as prior
  from weekly_activity
  where week_start >= current_date - 56 and week_start < current_date
  group by 1)
select a.account_id from accounts a join w using (account_id)
where a.ended_at is null and a.cancel_requested_at is null
  and current_date - a.signup_date >= 90
  and w.recent <= 0.60 * w.prior;
```

D4 · INTERVENTION + COMPARISON GROUP TEST SPEC

Test one thing on R1, against a comparison group

Item	Pre-registered choice
Intervention	A personal check-in from customer success at day 14–21: a 20-minute setup call focused on inviting the first teammate. No discount.
Mechanism	Accounts with one user never reach the shared workflows the product is priced for. Getting a second user in is the lever; the call is the delivery.
Segment / eligibility	R1: Monthly plan, signed up self-serve, no second user invited in the first 14 days, under 90 days old. Aged 14–89 days at test start. 196 accounts qualify on 01 Oct 2026.
Randomization unit	Account. Seeded random assignment by us; the client's team works the treatment list only.
Split	50/50: 98 treatment, 98 comparison group. The comparison group gets business as usual, nothing extra.
Duration	8 weeks from enrollment (single enrollment at test start (stock of matching accounts)).
Primary metric	Share of accounts that request cancellation within 56 days. Logo-based; ARR-weighted reported as secondary.
Baseline	14.7%: 56-day rate for R1 accounts aged 14–89 days across the lookback (3,887 account-snapshots).
MDE at 80% power	11.0 pts (from 14.7% to 3.7%), a 75% relative drop. Two-sided $\alpha = 0.05$.
Decision rule	Lift detected if the 95% interval for (treatment – comparison group) is entirely below zero. No lift if the interval is entirely above zero or sits within ± 2 pts. Anything else is directional only. Logo metric decides; ARR does not override it.

Said before the test starts

- At 98 per arm this test can only confirm a very large effect. To detect a one-third drop (to 9.8%) it needs about 695 accounts per arm. New R1 accounts arrive at about 22 a week, so rolling enrollment would take roughly 55 more weeks.
- Recommendation: run it anyway as a directional test (the intervention is cheap and not a discount), and say now that the readout will most likely be directional. This matches the Offer's rule: under ~400 eligible accounts, D7 is directional unless we extend or pool cycles.

D5 · EXECUTIVE BRIEF (EXCERPT)

One page for the person who signs off

What we looked at

Billing history for 8,000 accounts over two years and a weekly activity count. We agreed one churn number first (D1: 27.6% of accounts over the last 12 months, counted at end of term).

What we found (in the made-up data)

- Churn is not spread evenly. Monthly accounts without a second user by day 14 are 37% of signups and 62% of all accounts that ended.
- Four groups you can name with a rule carry 2.2–3.7× the base 90-day churn rate (D3). Together they hold \$658,841 of expected ARR at risk over the next 90 days (the segments overlap, so this sum counts some accounts twice).
- Annual customers decide at renewal. Their risk shows up in the 45 days before the renewal date, not mid-term.

The one recommendation

Test a setup call that gets a second user into new single-user monthly accounts (R1). It costs about 0.5 CS hours per account and gives nothing away.

What we are testing, and when you will know

98 accounts get the call and 98 are kept as the comparison group. Readout 8 weeks after the first calls. Be ready for a directional answer: the group is small enough that only an effect of 11 points or more would be conclusive (D4).

What we are not saying

- That these segments cause churn. They identify where to act; the test says whether acting helps.
- That any number here applies to another company. (And in this sample, none of it is real.)

D6 · AUDIENCE EXPORT

The list your team works from

Segment R1 as a CSV of account IDs with their arm, plus the definition in SQL so it can be rebuilt in the tool of record. 196 rows; first 12 shown. All IDs are synthetic (ACC-xxxxx). The full sample CSV ships next to this PDF.

account_id	segment_rule	plan	arr	tenure_days	arm
ACC-00025	R1	Starter (monthly)	\$588	88	comparison
ACC-00085	R1	Starter (monthly)	\$588	83	comparison
ACC-00095	R1	Starter (monthly)	\$588	71	treatment
ACC-00167	R1	Starter (monthly)	\$588	35	treatment
ACC-00252	R1	Starter (monthly)	\$588	51	comparison
ACC-00283	R1	Starter (monthly)	\$588	68	comparison
ACC-00310	R1	Starter (monthly)	\$588	88	comparison
ACC-00349	R1	Starter (monthly)	\$588	32	comparison
ACC-00365	R1	Starter (monthly)	\$588	20	treatment
ACC-00389	R1	Starter (monthly)	\$588	50	treatment
ACC-00480	R1	Starter (monthly)	\$588	47	treatment
ACC-00486	R1	Starter (monthly)	\$588	79	comparison

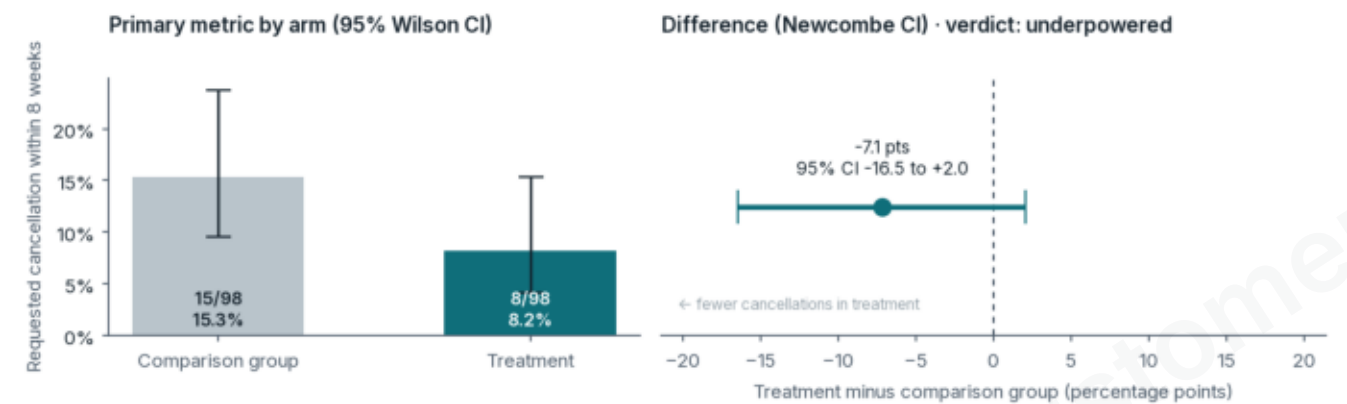
```
-- D3 / D6 · R1 early-tenure cliff (generic schema; adapt names to the client's warehouse)
select a.account_id, a.plan, a.arr, current_date - a.signup_date as tenure_days
from accounts a
where a.billing = 'monthly'
      and a.channel = 'self-serve'
      and a.cancel_requested_at is null
      and current_date - a.signup_date between 14 and 89
      and not exists (select 1 from users u
                      where u.account_id = a.account_id
                           and u.is_owner = false
                           and u.invited_at < a.signup_date + interval '14 days');
```

- The client's team contacts treatment accounts only. Comparison group accounts stay on business as usual; a monthly 15-minute check confirms nobody "helped" them.
- Arms are assigned once, by seed, before contact starts. Accounts are not re-drawn.

D7 · LIFT READOUT (SIMULATED)

Verdict: underpowered — directional only

SYNTHETIC SAMPLE
D7 · Lift readout vs. comparison group (synthetic, simulated outcome)



SYNTHETIC SAMPLE — not customer data · seed 20261020 · illustrative only, not a finding or benchmark

Measure	Treatment	Comparison group	Difference (95% CI)
Requested cancellation in 8 weeks (primary)	8/98 = 8.2%	15/98 = 15.3%	-7.1 pts (-16.5 to +2.0); p = 0.13
ARR-weighted (secondary)	6.5%	15.7%	-9.1 pts (no interval; higher variance)
Accounts kept, treatment arm	7.0		-2.0 to 16.1
ARR kept (annual value)	\$6,473		-\$1,856 to \$14,911
Cost of the calls	\$3,675		98 × 0.5 h × \$75/h (assumed)
Net of cost	\$2,798		-\$5,531 to \$11,236

Plain reading

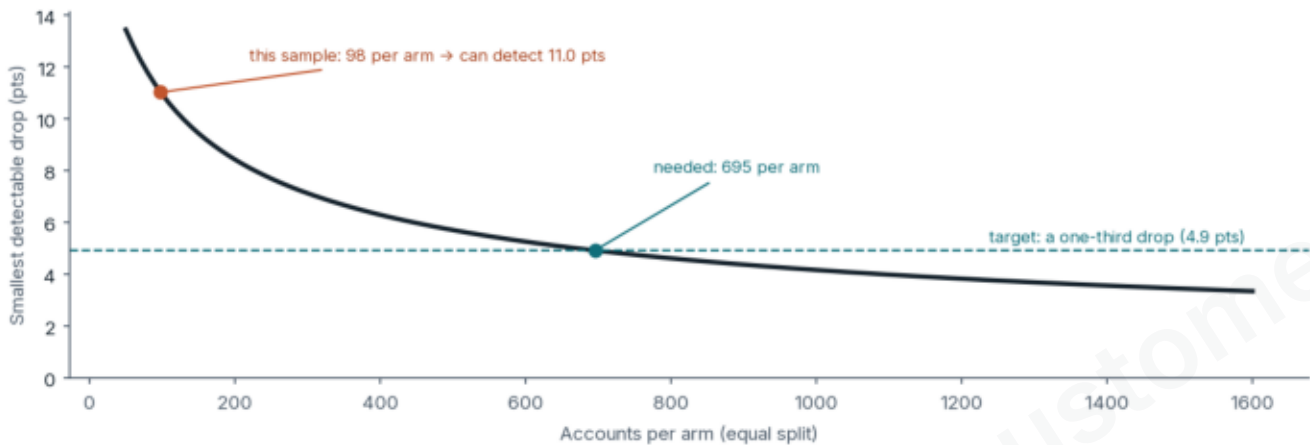
- Fewer treatment accounts asked to cancel (8.2% vs. 15.3%), but the interval runs from a large benefit to a small harm. The test was sized as directional (D4) and that is what it returned. This is not evidence that the calls work, and not evidence that they don't.
- Next step options: keep calling R1 accounts and pool a second 8-week cycle into the same analysis (pre-registered now), or stop if the CS hours are needed elsewhere.
- Simulation note: outcomes on this page were drawn at random with a built-in 25% relative reduction for treatment, from the D4 baseline. A real readout uses observed outcomes and has no such setting.

POWER

How big a test has to be

SYNTHETIC SAMPLE

Power · minimum detectable effect, baseline 14.7%, $\alpha = 0.05$, power = 80% (synthetic)



SYNTHETIC SAMPLE — not customer data · seed 20261020 · illustrative only, not a finding or benchmark

Power is the chance a test detects an effect that is really there. We plan for 80%, with a 5% false-alarm rate (two-sided). The size needed depends on how common the outcome is and how big a change you need to see. Sizes below are per arm, equal split, from statsmodels (Cohen's h).

Scenario	Baseline	Target	Accounts per arm
Whole base	3.00%	2.25%	7,097
At-risk segment	30.00%	20.00%	292
This sample's R1 test (one-third drop)	14.7%	9.8%	695

- Testing the whole base for a small monthly change needs about 7,097 accounts per arm, more than most companies this size have. That is why the Diagnostic tests a concentrated segment.
- A high-risk segment with a big expected effect needs far fewer (about 292 per arm).
- This sample sits in between: 196 eligible accounts, so the readout is directional unless enrollment is extended or cycles are pooled. We say that in the SOW, before anyone pays.

METHODS AND REPRODUCIBILITY

How this sample was made

Data

- generate_data.py invents 8,000 accounts with random signup dates, plans, prices, channel and whether a second user joined by day 14, plus a weekly count of core actions. Seed 20261020.
- It builds in exactly three patterns, the v1 segment hypotheses: (1) an early-tenure cliff for monthly accounts with no second user, (2) annual accounts leaving at renewal, (3) activity falling over the ~6 weeks before a cancellation request, plus random activity dips in healthy accounts so the activity rules have false positives.
- Because the patterns are built in, the analysis recovering them is circular. The pack shows what the deliverables look like, not that the method finds anything in real data.

Analysis

- D1: three trailing-12-month definitions on the same accounts. D2: Kaplan-Meier (lifelines) with 95% CI bands; multivariate log-rank test. Hazard by 30-day bin (bins with fewer than 50 at risk skipped).
- D3: monthly snapshots; outcome = ended within 90 days; stability checked in two halves split at Oct 2025.
- D4 / power: statsmodels NormalIndPower with Cohen's h. D7: Newcombe interval for the difference in proportions, Wilson intervals per arm, two-sample z-test p-value.
- Parameters that are assumptions, not data: target effect one-third; CS cost 0.5 h per account at \$75/h; simulated true effect 25%.

Rebuild it

```
cd sample-pack
python -m venv .venv && . .venv/bin/activate
pip install -r requirements.txt
python build_pack.py      # same seed -> same numbers, charts and PDF
```

Limits

- No real company, client or customer is described or implied. Plan names and prices are invented.
- Real engagements use the client's own definitions and data and usually find messier, smaller effects.