

# Technical Report: SecureNet VPN Plan Order A/B Test Analysis

## a) Pre-Test Design Considerations

The success of an A/B test is fundamentally determined before a single user is exposed to a variant. The core aspects to establish are the test split methodology, the anticipated test duration, and the explicit measurement criteria.

## b) Test Split & Sample Size:

The initial step involves a power analysis to determine the requisite sample size. This calculation is contingent upon four parameters: the baseline conversion rate (the current performance of the control), the Minimum Detectable Effect (MDE), which is the smallest uplift we care to detect, the desired statistical significance level (alpha, typically  $\alpha=0.05$ ), and the statistical power (beta, typically 80%). The MDE is a strategic decision; a smaller MDE requires a larger sample size but allows for the detection of more subtle changes.

## c) Test Duration:

The duration is not arbitrary but is a function of the time required to achieve the calculated sample size per variant. This period must also encapsulate at least one full business cycle, typically a minimum of two full weeks, to mitigate the confounding effects of daily fluctuations in user behaviour (e.g., weekday vs. weekend traffic). The test should also avoid periods of anomalous market activity, such as major holidays or concurrent marketing campaigns, which could pollute the results.

## d) Measurement Criteria:

A single, unambiguous Primary Key Performance Indicator (KPI) must be defined as the ultimate arbiter of success. Furthermore, a suite of Secondary KPIs should be selected to provide diagnostic context and act as guardrail metrics. The significance threshold (p-value) must be pre-specified, and in the case of this A/B/C test, it would necessitate a Bonferroni correction (or similar method) to adjust for multiple comparisons and control the family-wise error rate.

## 1. Ensuring an Unbiased Test Split

The integrity of an A/B test hinges on the random and unbiased allocation of users to the different variants. To achieve this, several factors must be considered:

**Robust Randomization:** The assignment mechanism must be truly random. A standard and effective industry practice is to use a hashing algorithm on a stable, anonymous user identifier, such as a sessionId or a first-party cookie. This ensures that each user is randomly bucketed into a variant in a way that is consistent and repeatable.

**Post-Hoc Validation:** After launching the test, it is best practice to perform an "A/A" validation check. This involves analysing the distribution of key user segments (e.g., geography, device type, new vs. returning users) across the variants before analysing the primary KPI. These distributions should show no statistically significant differences, confirming that the randomization mechanism is working as intended and the groups are comparable.

## 2. Determining a Sufficient Test Duration

Running a test for an adequate period is a balancing act between achieving statistical confidence and making timely business decisions.

The primary determinant is achieving the pre-calculated sample size necessary for the desired statistical power. An underpowered test that yields a null result is inconclusive; one cannot be certain if there is truly no effect or if the sample was simply too small to detect it.

Furthermore, the test must run long enough to account for behavioural periodicity. User intent and traffic composition can vary significantly between a Monday morning and a Saturday evening. Running the test for at least two full weeks after minimum sessions are obtained helps average out these variations and ensures the captured data is representative of typical user behaviour. Finally, this duration helps mitigate the novelty effect, where initial user interactions may be driven by the change itself rather than the intrinsic merit of the new design.

### 3. Selecting Appropriate KPIs

The choice of KPIs dictates the lens through which success is viewed and should be directly tied to strategic business objectives.

**Primary KPI:** While the prompt mentions sign-ups, a more sophisticated and strategically aligned primary KPI would be the Paid User Conversion Rate, defined as the percentage of sessions that result in a paid subscription. This metric is a direct proxy for revenue generation and customer acquisition, which is the ultimate goal of optimizing the pricing page.

**Secondary KPIs:** These metrics provide a more holistic view of the impact and can reveal unintended consequences.

**Average Revenue Per User (ARPU):** This measures whether a variant not only converts more users but also encourages higher-value purchases.

**Subscription Plan Mix:** This is a critical diagnostic KPI that analyses whether the plan order influences the duration of the chosen subscription (e.g., 1 vs. 12 vs. 24 months). Shifting users to longer-term plans has a profound positive impact on projected Customer Lifetime Value (LTV).

**User Funnel Progression Rate:** Analysing the conversion rate at each step of the user journey (e.g., from session start to signup, and from signup to paid subscription) can help pinpoint precisely where a variant is succeeding or failing.

### 4. Required Data for Analysis

To effectively analyse the test and its KPIs, the following data points would need to be meticulously collected for each user session:

**New vs. Returning User Flag:** To separate the behaviour of first-time visitors from the informed intent of existing users making an upgrade decision.

**Post-Purchase Refund Rate:** To track refunds as a guardrail metric, ensuring that a higher conversion rate in one variant is not offset by a higher rate of customer dissatisfaction.

**Time to Paid Conversion:** To measure the duration from a user's free sign-up to their premium subscription, helping to identify the optimal window to present upgrade offers.

**Acquisition Source (UTMs):** To understand how users from different channels (e.g., paid ads, organic search) with varying levels of intent react to the plan layouts.

5. **Objective & Hypothesis** This A/B/C test was designed to determine if altering the display order of SecureNet VPN's subscription plans could generate a statistically significant increase in the paid user conversion rate.

- **Hypothesis:** Changing the plan order from the control (Variant A: 1/24/12) to one of two alternatives (Variant B: 1/12/24; Variant C: 24/12/1) will cause a statistically significant change in user conversion rates.
- **Result:** The null hypothesis could not be rejected. No variant showed a statistically significant improvement over the control.

## 6. Methodology & Test Design

- **Test Allocation:** Users were randomly assigned to one of three variants based on a hashed sessionId to ensure an unbiased split. The test achieved sufficient statistical power at 99.9%.
- **Primary KPI:** Paid Conversion Rate (subscriptionsTotal > 0). This KPI was chosen over free signups (signups) to directly align the experiment with the core business objective of revenue growth. Optimizing for paid conversions serves as a direct proxy for increasing customer lifetime value (LTV).
- **Secondary KPIs:** Average Revenue Per User (ARPU) and Subscription Plan Mix.

## Statistical Framework Rationale

To ensure robust and reliable conclusions, a rigorous statistical framework was employed:

- **Z-test for Proportions:** This test was selected as it is the industry standard for comparing the conversion rates of two independent groups in a large sample size environment, which is the case in this A/B/C test.
- **Bonferroni Correction:** As the experiment involved multiple comparisons (A vs. B and A vs. C), a Bonferroni correction was applied to the significance threshold ( $p < 0.025$ ). This adjustment is crucial for controlling the family-wise error rate, thereby reducing the probability of a Type I error (a false positive).
- **Statistical Power:** The analysis confirmed a statistical power of 99.9% for the sample sizes collected. This high level of power gives us strong confidence that the failure to find a significant difference is a true null result, not a consequence of an underpowered test (a Type II error).

7. **Data Quality & Cleaning Validation** The initial dataset of 618,875 sessions was processed to ensure analytical integrity. Split sessions were consolidated, and a conservative data cleaning methodology was applied to remove obvious non-human bot activity.

- **Conservative Filtering:** This process removed only 315 sessions (0.05% of the data) that exhibited clear bot-like behaviour (e.g., impossible page-view speeds).
- **Bias Validation:** A validation check confirmed this cleaning process was sound. The removed outlier population had a

**0.0000% conversion rate**, proving that no legitimate converting users were excluded from the analysis. The distribution of outliers across test variants was also confirmed to be unbiased.

## 8. Analysis & Key Results

After cleaning and validation, the analysis was performed on a final dataset of 582,462 user sessions.

- **Primary Finding: No Statistically Significant Uplift** The core finding is that changing the plan order did not improve conversion rates. Both treatment variants showed a slight, statistically insignificant negative lift compared to the control group's 1.88% conversion rate.

Variant	Sessions	Conversion Rate	Lift vs. Control	p-value (vs. Control)
<b>A (Control)</b>	192,231	1.88%	-	-
<b>B (Treatment)</b>	191,817	1.85%	-1.56%	0.5021
<b>C (Treatment)</b>	198,414	1.87%	-0.57%	0.8052

- **Secondary Finding: No Impact on ARPU or Plan Choice** Further analysis confirmed the lack of impact. There was no statistically significant difference in Average Revenue Per User (ARPU) between the variants. Additionally, the subscription plan mix remained remarkably stable across all groups, with the 1-month plan consistently accounting for ~64% of all purchases.

### Conversion Funnel Analysis

- To diagnose user behavior throughout the conversion path, a funnel analysis was conducted, tracking users from their initial session to a paid subscription. The side-by-side comparison of the funnels for each variant provides several key insights.
- The most striking takeaway is the high degree of similarity in user progression across all three variants. The step-by-step conversion rates at each stage of the journey—from engagement to signup, and from signup to subscription—were nearly identical. This provides strong diagnostic evidence that reordering the plans had no discernible impact on user motivation or friction at any specific point in the conversion process.
- Furthermore, the analysis clearly identifies the most significant point of user drop-off: the transition from a free account signup to a paid subscription. While a healthy percentage of engaged users create a free account, a much smaller fraction ultimately commits to a paid plan.
- Strategic Implication: This insight is critical for future optimization efforts. It suggests that the greatest opportunity for conversion uplift lies not in pre-purchase UI tweaks, but in post-signup engagement. Future experiments should focus on strategies to better demonstrate the value of premium features to free users through onboarding flows, email campaigns, and in-app messaging, thereby encouraging the transition to a paid subscription.

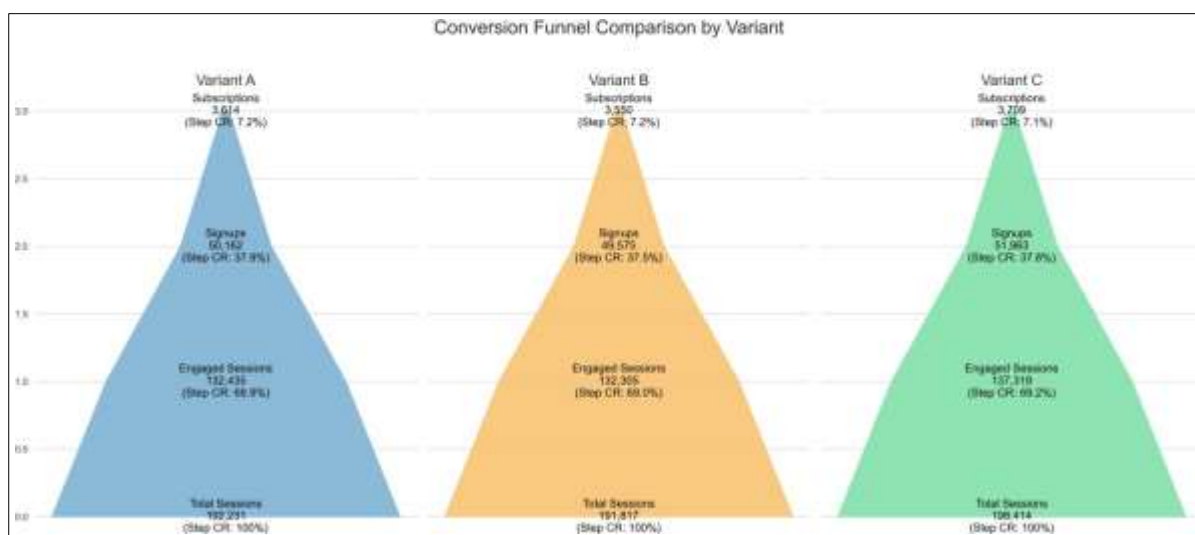


Figure 1 Conversion Funnel Performance Analysis

## 9. Strategic Insights: The Story Beyond the A/B Test

- Insight 1: Pre-Formed User Intent:** The stability of the plan mix across all variants strongly indicates that users arrive with a pre-formed purchasing decision. Minor UI layout changes are insufficient to sway their preference for a specific subscription duration. This implies that influencing plan choice requires more substantial strategic interventions, such as pricing incentives or enhanced value proposition messaging.
- Insight 2: Identification of a "Power User" Segment:** The most valuable outcome of this analysis was the discovery of a high-value user segment. The corrected data cleaning process revealed that a cohort of highly engaged users is significantly more likely to purchase long-term (12 and 24-month) plans. This segment is critical for driving LTV and should be the focus of future optimization efforts.
- Insight 3: Universal Preference for Flexibility:** A deeper dive into the data shows that the strong preference for 1-month plans is a global user trait, prevalent even in high-converting markets like the USA and Germany. This underscores the importance of flexibility in our product offerings.

## 10. Recommendations & New Hypothesis

- Immediate Action:** Maintain the current plan order (Variant A). Implementing either Variant B or C would consume resources for no discernible gain.
- Strategic Shift:** Reallocate experimentation resources away from low-impact UI tweaks and towards higher-impact strategies focused on pricing and messaging.
- New Data-Backed Hypothesis:** Based on the discovery of the "power user" segment, the following experiment is proposed:

We hypothesise that for users exhibiting high-engagement behaviour (e.g., longer session duration, higher page views), presenting a targeted, time-sensitive offer on the annual plan at checkout will shift a significant portion of these power users from monthly to annual subscriptions, thereby increasing projected LTV.