# Elements of Permutation Test with Application to Vaccine Study

Lei Li, PhD

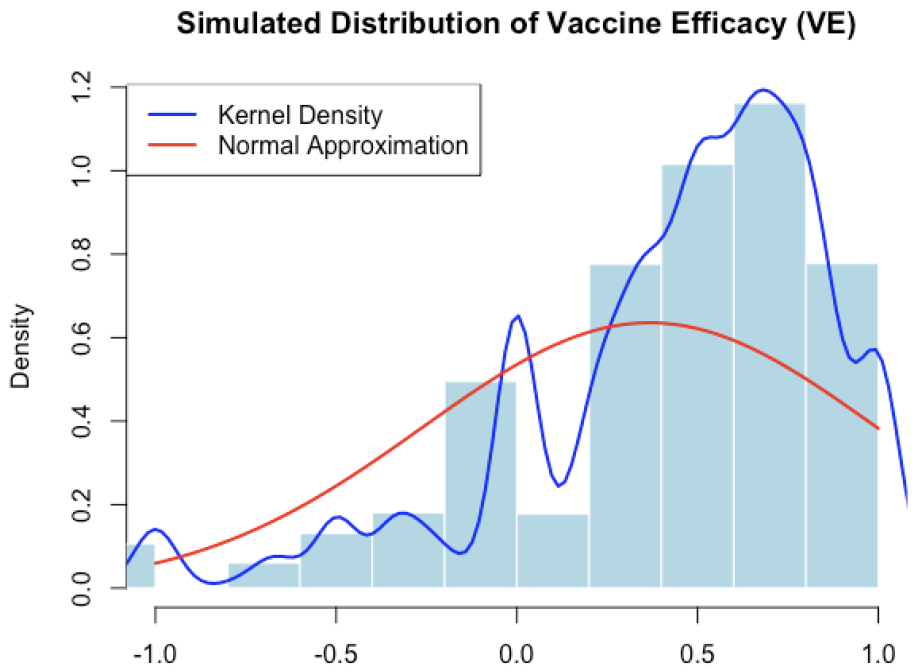
July 15, 2025

# Agenda

- Motivation Example
- Introduction to Permutation Test (PT)
- p value and power calculation based on PT
- Theoretical Properties of Studentized Permutation Test
- Simulation Studies
- Concluding Remarks

# Motivating Example

- **Vaccine Efficacy (VE)** is defined as $VE = 1 - \frac{p_1}{p_0}$, where $p_1$ and $p_0$ denote the attacking rate from vaccine arm and placebo arm, respectively.
- However, VE is generally **not normally distributed**, especially with small sample sizes, leading to non-robust inference.
- Early-phase trials (e.g., phase I/II) often involve limited sample sizes.
- **Normal-theory methods** (e.g., Wald type method) may give inaccurate inference (e.g. type I error inflation).



# Permutation Test and Example

- A **permutation test** is a non-parametric exact hypothesis test.
- **Hypothesis Testing**: $H_0$: F = G vs $H_a$: F $\neq$ G.
- Under $H_0$, the distribution of the test statistic is obtained by calculating all possible values of the test statistic under possible rearrangements of the observed data.
- This is a model free method and the incurred inference is exact inference.

**Example:** Group 1: 1, 2 vs Group 2: 3, 4

- Observed Mean Difference $-2$.
- Total shuffles: $\binom{4}{2} = 6$.
- Permutation p-value: the frequency of test statistic as extreme as or more than the observed value: $1/6 = 0.17$.
- t-test p-value: 0.11.

## Permutation Test Example

| Group 1 | Group 2 | Diff |
|---|---|---|
| 1, 2 | 3, 4 | $-2$ |
| 1, 3 | 2, 4 | $-1$ |
| 1, 4 | 2, 3 | 0 |
| 2, 3 | 1, 4 | 0 |
| 2, 4 | 1, 3 | 1 |
| 3, 4 | 1, 2 | 2 |

# Permutation Test Conditions

- For a 1:1 allocation, there are $C_{2n}^{n}$ combinations, where $n$ is the sample size per group.
- For large $n$, simulating all permutations is infeasible, e.g., $C_{100}^{50} \approx 10^{29}$.
- **Solution:** Use partial permutation samples to approximate the distribution.
- **Key advantage of PT:** No assumptions on distribution $\Rightarrow$ robust to misspecification.

Table: Permutation Test vs Parametric Test

| **Feature** | **Permutation Test** | **Parametric Test** |
|---|---|---|
| **Purpose** | Compare groups non-parametrically under unknown or violated assumptions | Compare group means assuming known distributions |
| **Assumption** | No assumptions about data distribution; robust to normality/variance violations | Assumes normality and equal variances |
| **Test Statistics** | From permutation of data | Student t-test, F-test, etc. |

# Permutation Test p-value and Conditional Power

**Calculate (conditional) p-value:**

**Algorithm 1: Conditional Monte Carlo (CMC)**

- Step 1: Given dataset $D$ (e.g., vaccine vs placebo), compute test statistic $T_0 = T(D)$.
- Step 2: Permute $D$ randomly to get $D^*$, compute $T^* = T(D^*)$.
- Step 3: Repeat Step 2 $B$ times to get $T_1^*, \ldots, T_B^*$.
- Step 4: Compute the (conditional) p-value:

$$\hat{p}_B(D) = \frac{1}{B} \sum_{b=1}^{B} \mathbb{I}(T_b^* \geq T_0)$$

**Statistical Properties:**

- $\hat{p}_B(D)$ is an unbiased estimator of $p(D)$ by Glivenko–Cantelli theorem.
- Confidence interval of $\hat{p}_B(D)$:

$$\hat{p}_B(D) \pm z_{\alpha/2} \sqrt{\frac{\hat{p}_B(D)(1 - \hat{p}_B(D))}{B}}.$$

**Calculate (conditional) power**

Recall the definition of conditional power:

$$1 - \beta(D') = P(\text{reject } H_0 \mid H_a, D')$$

**Algorithm 2: Conditional Power Algorithm**

- Step 1: For dataset $D'$, compute $\hat{p}_B(D')$ using Algorithm 1.
- Step 2: Repeat for $R$ permutations to get $\hat{p}_B(D'_1), \ldots, \hat{p}_B(D'_R)$.
- Step 3: Estimate conditional power:

$$1 - \hat{\beta}(D') = \frac{1}{R} \sum_{r=1}^{R} \mathbb{I}(\hat{p}_B(D'_r) \leq \alpha)$$

**Note:** Conditional power can guide interim analysis (e.g., futility stopping).

# Studentized Permutation Test

- The traditional permutation test requires the assumption of samples from two groups are exchangeable (e.g., applied to bioequivalence study), however, in vaccine efficacy, the hypothesis usually is a superiority test, e.g., $H_0 : VE \leq 0.2$ vs $H_a : VE > 0.2$, violating the exchangeability assumption. One solution is to use **studentized permutation test**.

# Permutation limit for the studentized VE statistic

## Theorem (Studentized Permutation limit)

*Let* $X_1, \ldots, X_{n_v} \overset{iid}{\sim} \mathrm{Ber}(p_v)$ *and* $Y_1, \ldots, Y_{n_c} \overset{iid}{\sim} \mathrm{Ber}(p_c)$ *be independent random variables across two arms. Define* $n = n_v + n_c$, *and*

$$\widehat{\mathrm{VE}} = 1 - \frac{\hat{p}_v}{\hat{p}_c}, \quad \text{where} \quad \hat{p}_v = \sum_{i=1}^{n_v} X_i \quad \text{and} \quad \hat{p}_c = \sum_{j=1}^{n_c} Y_j.$$

*Also, let the studentized permutation test (PT) statistic* $T_n$ *for VE be*

$$T_n = \frac{\widehat{\mathrm{VE}} - \mathrm{VE}_0}{\widehat{\mathrm{SE}}}, \quad \text{where} \quad \widehat{\mathrm{SE}} = \sqrt{\frac{\hat{p}_v(1-\hat{p}_v)}{n_v \hat{p}_c^2} + \frac{(1-\hat{p}_c)\hat{p}_v^2}{n_c \hat{p}_c^3}}.$$

*Additionally, denote* $T_n^\pi = T_n$ *as the studentized PT computed on a random label permutation. Assume* $p_c \in (\delta, 1-\delta)$ *for some* $\delta > 0$. *Then, under* $H_0 : VE = VE_0$ *agains* $H_a : VE \neq VE_0$, *the studentized permutation test statistic* $T_n^\pi$ *conditioning on the data* $(X_1, \ldots, X_{n_v}, Y_1, \ldots, Y_{n_c})$ *converges to a standard normal distribution, that is,*

$$\sup_{t \in \mathbb{R}} \left| \Pr\{ T_n^\pi \leq t \mid (\mathbf{X}, \mathbf{Y}) \} - \Phi(t) \right| \xrightarrow{P} 0, \ n \to \infty,$$

*i.e. the conditional permutation law of* $T_n^\pi$ *converges to* $N(0,1)$ *without any exchangeability requirement.*

# Valid test and CI derived from the permutation limit

Let $q^{\pi}_{1-\alpha}$ be the $(1-\alpha)$-quantile of $\left|T^{\pi}_{n}\right| \mid (\mathbf{X}, \mathbf{Y})$.

**Theorem (Test and CI)**

*For testing $H_0 : VE = VE_0$ vs $H_a : VE \neq VE_0$, then asymptotically, as $n = n_c + n_v \to \infty$,*
**(i) Studentized permutation test**

$$\varphi^{\pi}_{n} = \mathbf{1}\Big\{|T_n| > q^{\pi}_{1-\alpha}\Big\} \quad \textit{satisfies} \quad \Pr_{H_0}\{\varphi^{\pi}_{n} = 1\} \to \alpha, \quad \Pr_{H_a}\{\varphi^{\pi}_{n} = 1\} \to 1.$$

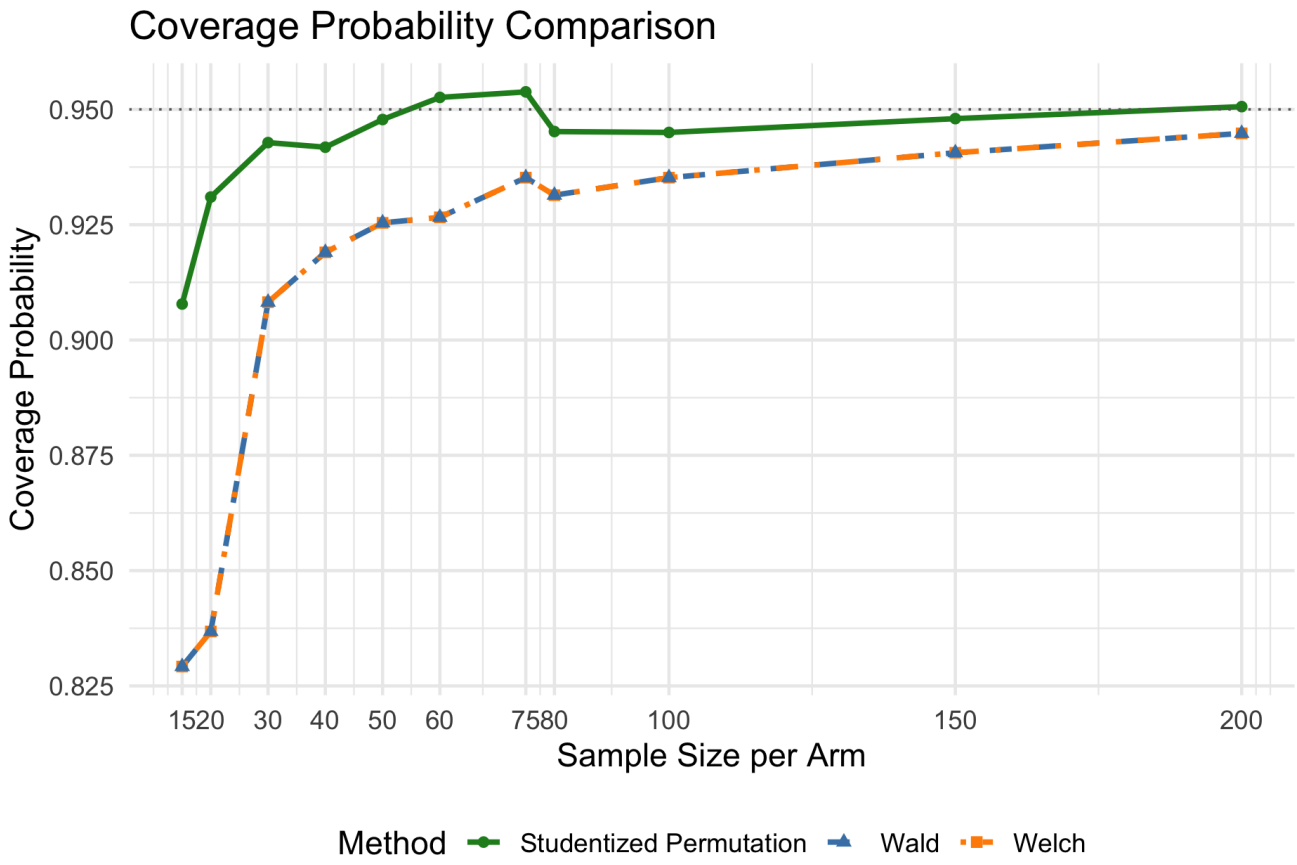
**(ii)** $(1-\alpha)$ **confidence interval for VE**

$$\mathcal{I}^{\pi}_{n} = \Big[1 - \exp\big(\log(1-\widehat{\mathrm{VE}}) + q^{\pi}_{1-\alpha}\widehat{\mathrm{SE}}\big),\ 1 - \exp\big(\log(1-\widehat{\mathrm{VE}}) - q^{\pi}_{1-\alpha}\widehat{\mathrm{SE}}\big)\Big]$$

*has asymptotic confidence level* $1-\alpha$*, that is,* $\Pr\{\mathrm{VE} \in \mathcal{I}^{\pi}_{n}\} \to 1-\alpha$.

# Simulation Study Coverage Probability

- **Two-sided Coverage Probability Setup**: Assume True $VE = 0.8$, attack rate from placebo arm $p_c = 0.4$. The simulation is repeated 5000 times.
- **Findings**: For small sample size, the studentized PT outperforms the other two methods.

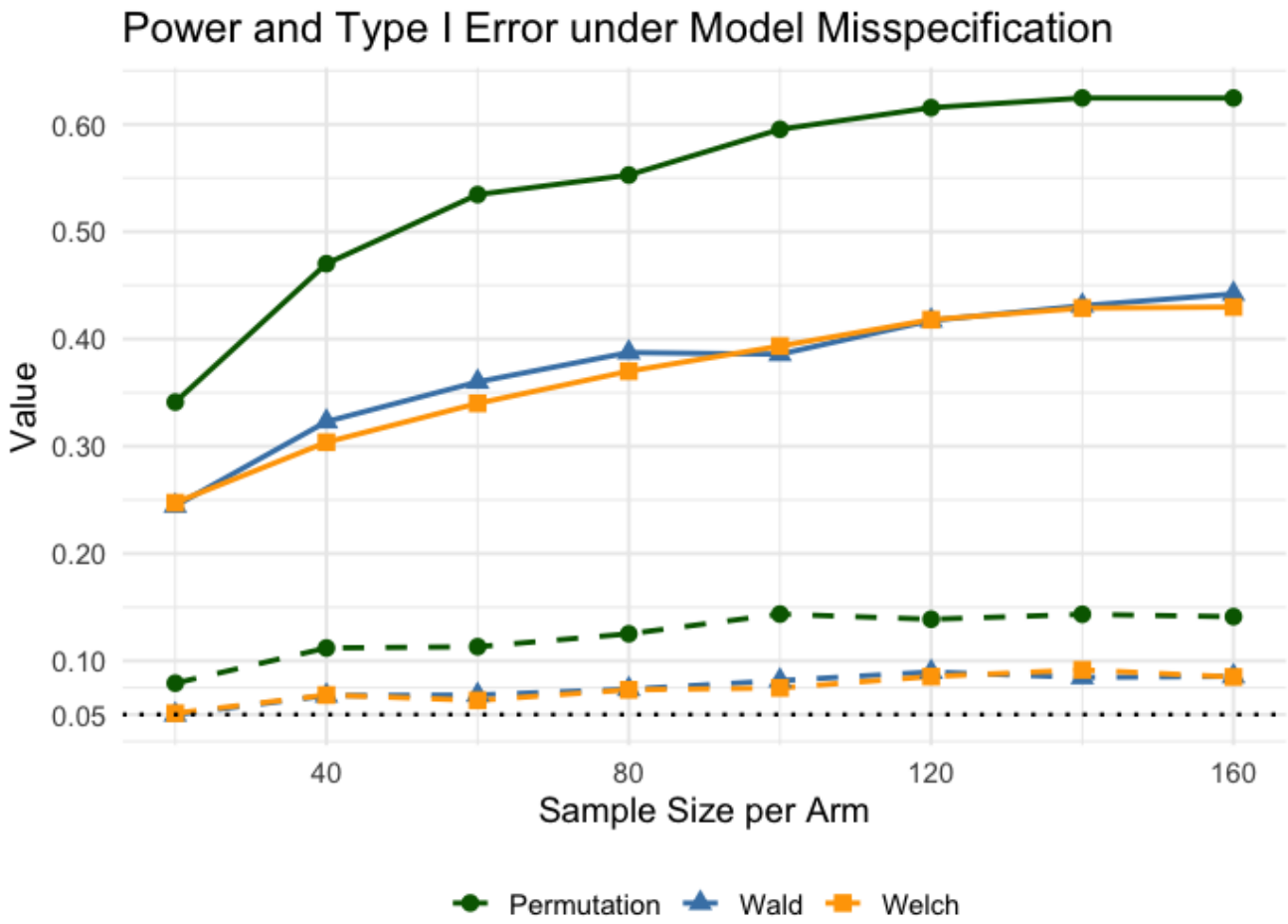


# Simulation Study Power and Type I error

- **Setup**: Assume $VE_0 = 0.4$, $VE_a = 0.85$, and $p_c = 0.4$.
- **Findings**: For the studentized permutation method, the type I error is well controlled even for small sample size, outperforming the other two commonly used methods. The power for both Wald and Welch method outperform studentized permutation method but also come at the cost of inflation of type I error.

# Simulation Study Power and Type I error with Model Misspecification

- **Setup**: We add random effects to attacking rates in both arms. Assume $VE_0 = 0.4$, $VE_a = 0.85$, $p_c \sim 0.8 * Beta(1, 1)$, and $p_v \sim 2 * Beta(1, 1) * (1 - VE_a)$.
- **Findings**: For SPT, the power performs much better than the other two methods, however, the type I error also is inflated.



# Studentized Permutation Test for Group Sequential Design

**Algorithm**: We implement Monte Carlo type SPT algorithm for sequential design: For $k$ looks, the overall type I error

$$P(VE_{t_1} > c_1 \text{ or } VE_{t_2} > c_2 \ \ldots, \text{or } \ldots, VE_{t_k} > c_k) = \alpha \tag{1}$$

is approximated by its Monte Carlo estimate

$$\frac{1}{B}\sum_{b=1}^{B} I(VE_{bt_1} > c_1 \text{ or } VE_{bt_2} > c_2 \ \ldots \text{or } \ldots VE_{bt_k} > c_k), \tag{2}$$

where $B$ is the number of Monte Carlo replications.

- For $k$ looks, calculate the studentized permutation test statistic $T_{t1}, T_{t2}, \ldots, T_{tk}$.
- For a candidate set of studentized permutation critical values $d_1, d_2, \ldots, d_k$ and a number of $B$ replications, approximate the overall type I error as $\alpha$ using Monte Carlo estimate, that is,
$$\frac{1}{B}\sum_{b=1}^{B} I(T_{bt_1} > d_1 \text{ or } T_{bt_2} > d_2 \ \ldots \text{or } \ldots T_{bt_k} > d_k) = \alpha.$$
- The critical values for $VE$ is given by $c_i = d_i \cdot SE + VE_0$ for $i = 1, \ldots, k$.

# Concluding Remarks

- The studentized permutation test enjoys both exact and asymptotic inference property: when exchangeability assumption holds, the permutation test inference is exact inference, not asymptotic; when the assumption does not hold, asymptotically it is also valid.
- The type I error is well controlled for small sample size.
- In sequential design setting, the studentized permutation test could yield sharper boundaries than normality theory based method.

# End

Questions Comments :)