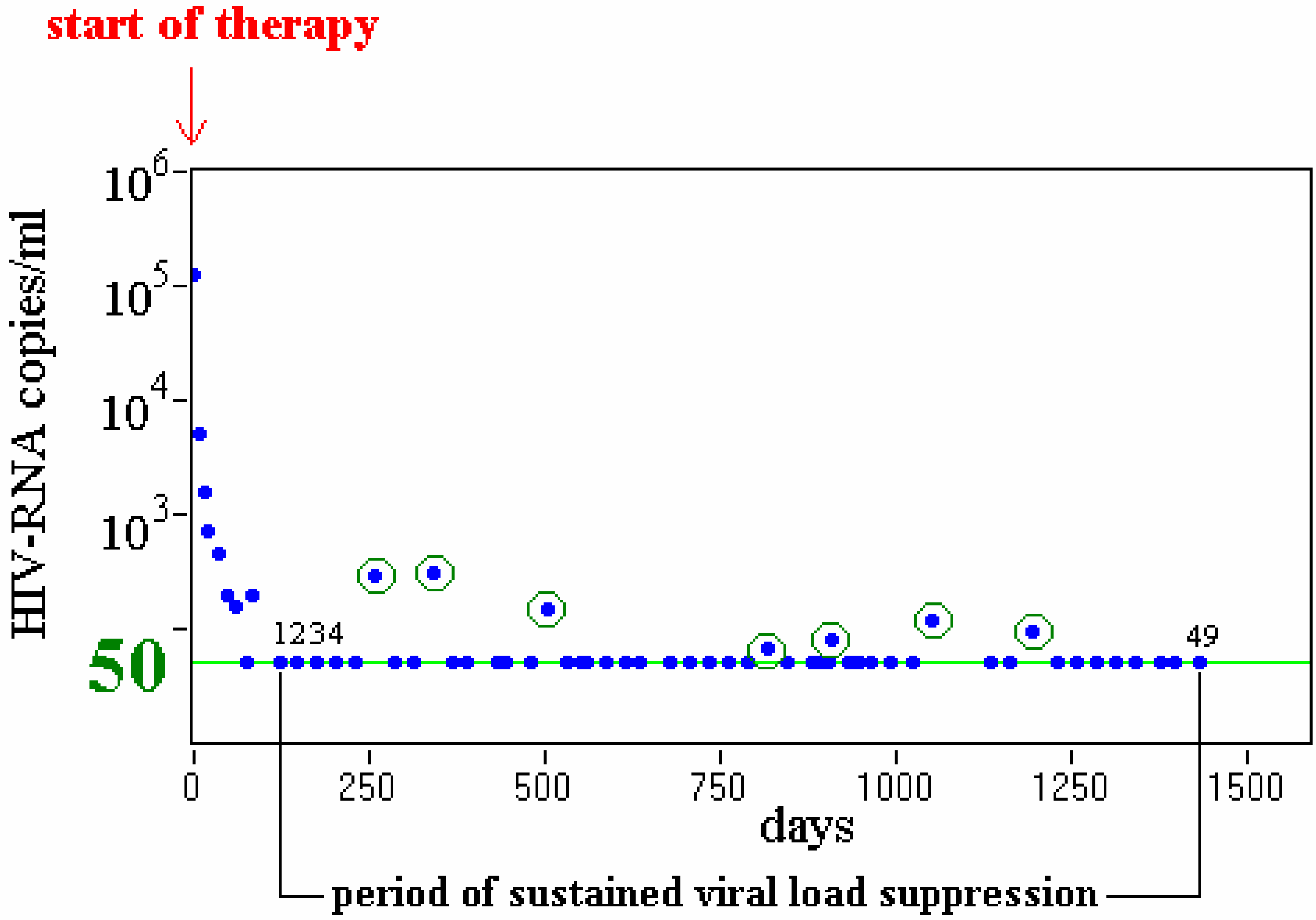
Multiple Outputation Permutation Exact Inference for Complex Clustered Data

Dean Follmann & Michael Fay

National Institute of Allergy and Infectious Diseases

Highly Active Antiretroviral Therapy (HAART)



Motivation

- Serial VL, CD4 measurements. What do blips mean?

Patient	Time	VL	CD4	blip
Joe	May 1 2004	<50	400	no
Joe	June 1 2004	1000	450	yes
...				
Joe	Sept 1 2004	<50	600	no
Joe	Oct 1 2004	<50	750	no

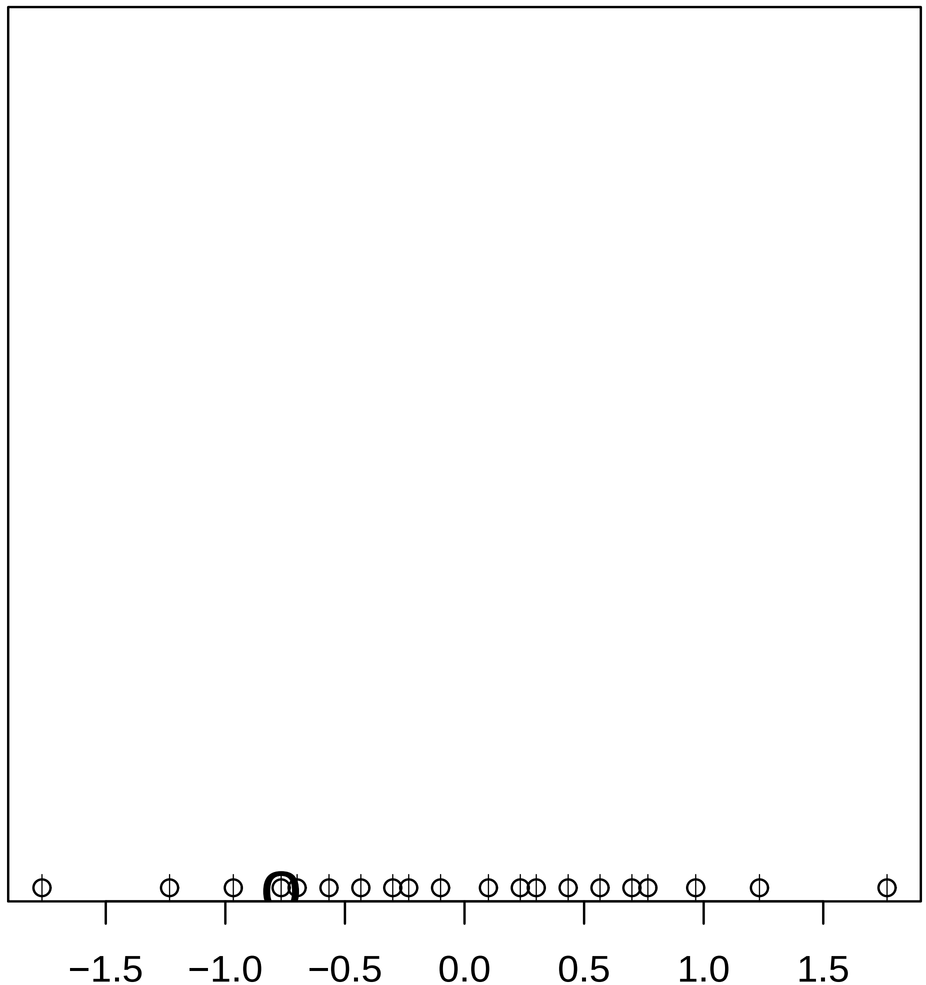
- $D = R_{01} - R_{00} = (450 - 400)/400 - (750 - 600)/600$ and do a paired-difference permutation t-test.
- But multiple nonblip-blip & blip-blip couples per person...

Permutation tests

Person	X	Z
1	1.1	0
2	1.5	0
3	2.3	0
4	3.3	1
5	3.1	1
6	.8	1

$$t_0 = t(\boldsymbol{X}, \boldsymbol{Z}) = \frac{\sum_{i=1}^{n} Z_i X_i}{3} - \frac{\sum_{i=1}^{n} (1 - Z_i) X_i}{3}.$$

Permutation Distribution for t_0



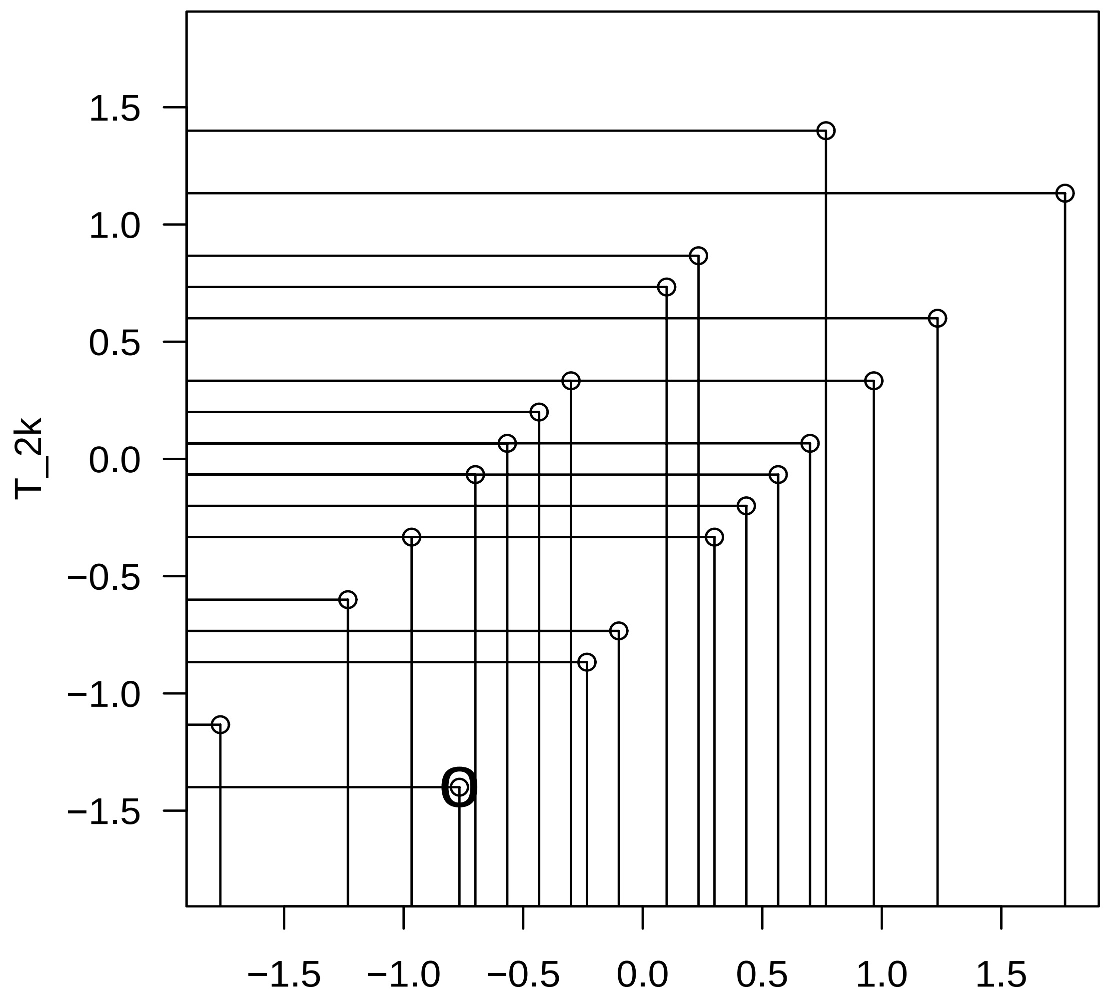
What about clustering?

Person	X	Z
1	1.1	0
2	1.5	0
3	2.3	0
4	3.3	1
5	3.1	1
6	.8 & 2.7	1

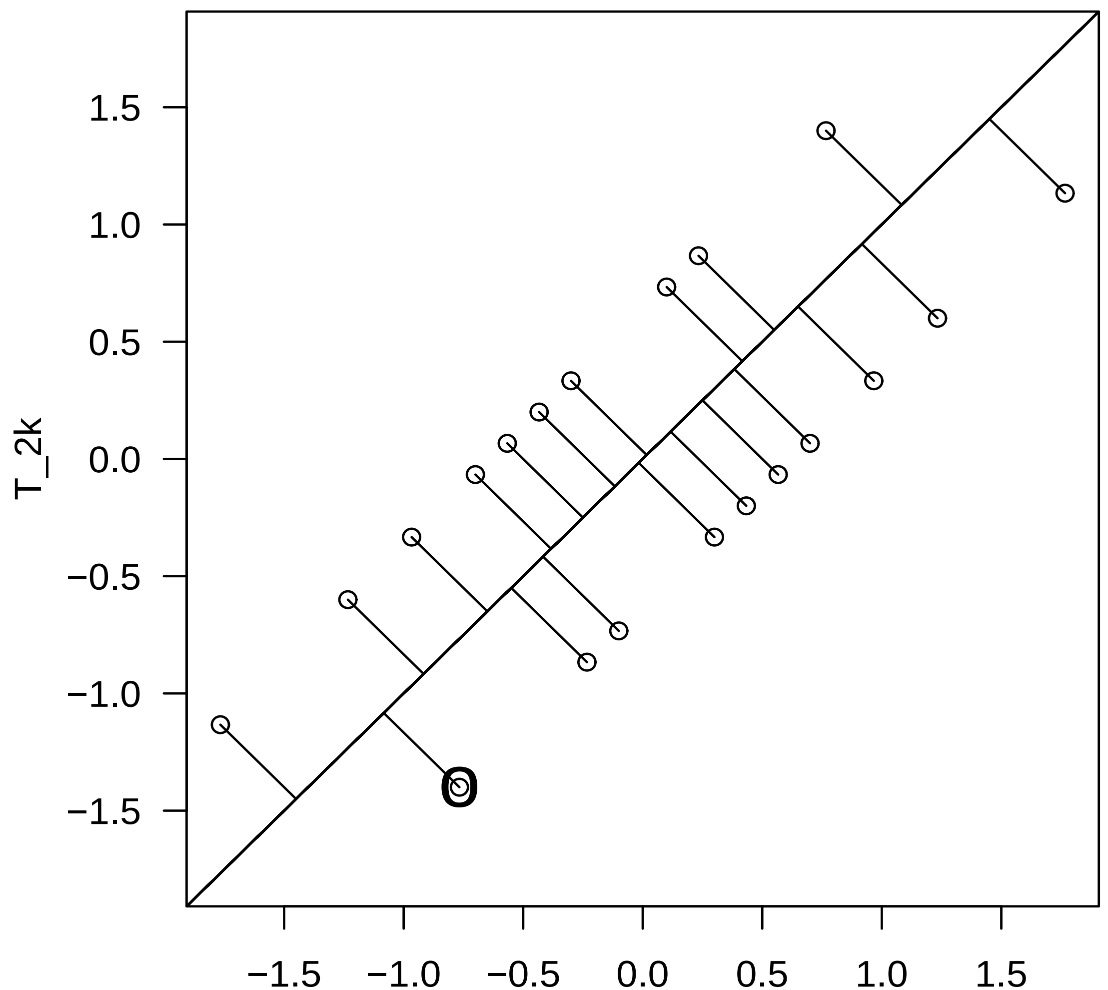
$$t_{j0} = t\{\boldsymbol{X}(\boldsymbol{j}_j), \boldsymbol{Z}\} = \frac{\sum_{i=1}^{n} Z_i X_i(\boldsymbol{j}_j)}{3} - \frac{\sum_{i=1}^{n} (1 - Z_i) X_i(\boldsymbol{j}_j)}{3},$$

$\boldsymbol{j}_j = (1, 1, 1, 1, 1, j)$ for $j = 1, 2$

Permutation distributions for t_{10} and t_{20}



Let's Combine



Exhaustive Outputation Permutation

- Fix a permutation of the Zs e.g. $\boldsymbol{\pi} = (8, 1, ..., 21)$.
- Throw *out* all data but one per cluster e.g. $\boldsymbol{j} = (1, 4, 1, ...3)$
- Form the test statistic $t\{(\boldsymbol{X}(\boldsymbol{j}), \boldsymbol{Z}(\boldsymbol{\pi})\}$
- Do this exhaustively for all $m = m_1 \times \ldots \times m_n$ outputations.
- Average over all outputations for this fixed $\boldsymbol{\pi}$.

$$\sum_{\boldsymbol{j}} \frac{t\{(\boldsymbol{X}(\boldsymbol{j}), \boldsymbol{Z}(\boldsymbol{\pi})\}}{m}$$

- The averaged test statistics form the permutation distribution.

Exhaustive Outputation Permutation

Matrix of test statistics for EOP.

	Permutation (k)				
Outputation (j)	$\boldsymbol{\pi}_0$	$\boldsymbol{\pi}_1$	$\boldsymbol{\pi}_2$	...	$\boldsymbol{\pi}_b$
$\boldsymbol{j}_1$	t_{10}	t_{11}	t_{12}	...	t_{1b}
$\boldsymbol{j}_2$	t_{20}	t_{21}	t_{22}	...	t_{2b}
.				t_{jk}	
$\boldsymbol{j}_m$	t_{m0}	t_{m1}	t_{m2}	...	t_{mb}
Average	$\overline{t_{.0}}$	$\overline{t_{.1}}$	$\overline{t_{.2}}$	...	$\overline{t_{.b}}$

Special Case of EOP: Permutation t-test

- For the permutation t-test, EOP is easy
  1. Form the within cluster means $\overline{X}_i$.
  2. Do usual permutation t-test on the $\overline{X}_i, Z_i$s.

Monte Carlo Outputation Permutation

- Note: EOP uses

$$y_k = I(\bar{t}_{\cdot k} - \bar{t}_{\cdot 0} > 0)$$

- $\bar{t}_{\cdot}$ based on m, use average $y_1, \ldots, y_b = \overline{y}$
- Randomly pick $M < m$ outputations $B < b$ permutations
- Choose M large enough so that

$$Y_K = I(\overline{T}_{\cdot K} - \overline{T}_{\cdot 0} > 0) \approx y_K$$

- Choose B large enough. Note

$$\sum_{K=1}^{B} \frac{Y_K}{B} \approx N(\overline{y}, \frac{\overline{y}(1-\overline{y})}{B})$$

Some Asymptotics

- We use $\overline{T}_{\cdot k}$ in lieu of $\bar{t}_{\cdot k}$

$$\overline{T}_{\cdot k} \approx N(\bar{t}_{\cdot k}, \frac{\sum_{j=1}^{m}(t_{jk} - t_{j0})^2}{m}$$

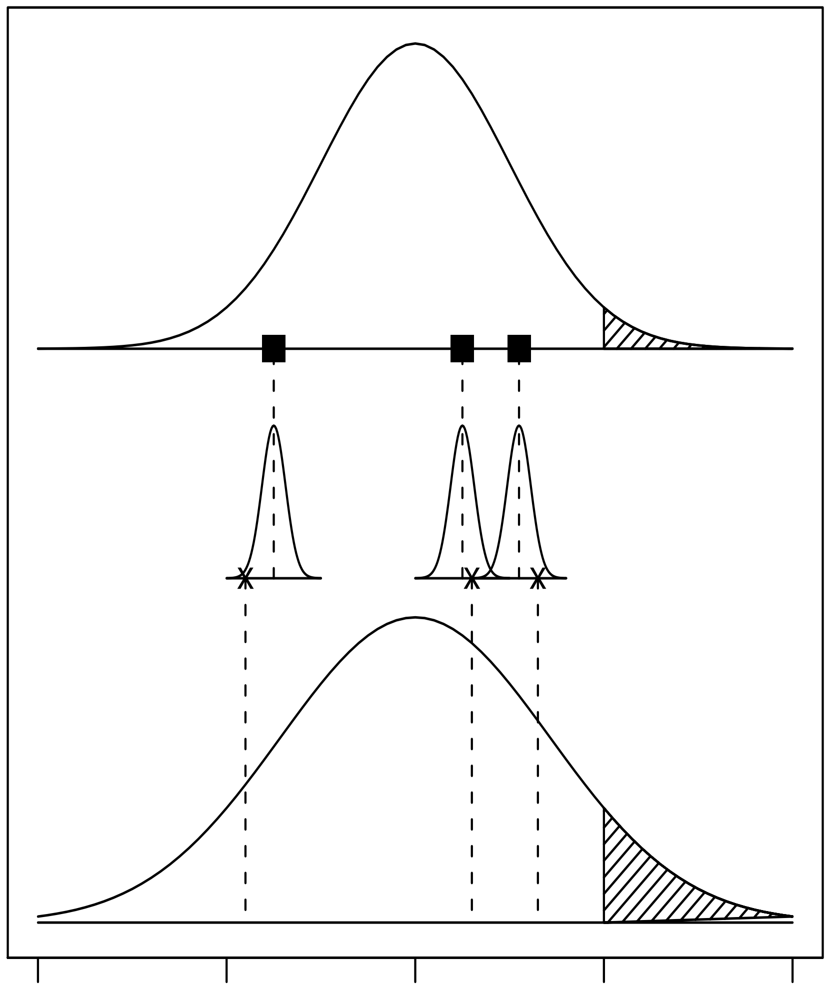
by the central limit theorem as $M \to \infty$

- Further, under certain conditions, we can invoke the permutation central limit theorem.

$$\bar{t}_{\cdot K} \approx N(0, \tau^2)$$

for the randomly selected permutation Π_K.

Asymptotic dbns for EOP & MOP



Informative Cluster Size

- Suppose that n randomized per group, sickest $n/2$ control patients are seen more often. Treatment has no effect.
-
$$H_0^{SN} : F_0(x|m) = F_1(x|m), \qquad p_0(m) = p_1(m)$$
$$H_0^{WN} : F_0(x|m) = F_1(x|m).$$
- EOP corresponds to a permutation t-test on within cluster averages and test H_0^{WN}
$$\frac{\sum \overline{X}_i Z_i}{\sum Z_i} - \frac{\sum \overline{X}_i (1 - Z_i)}{\sum (1 - Z_i)}$$

Example Revisited

- Is the relative change in CD4 same for nonblip-blip versus nonblip-nonblip couples?
- For each patient obtain all R_{01}s and R_{00} and form

$$\overline{D} = \overline{R_{01}} - \overline{R_{00}}$$

- Using these $\overline{D}$s do a permutation paired difference t-test.
- $n = 44$ with exact upper p-value = .1793.

Summary

- Outputation permutation simple idea. Make independent data by throwing data out, but do it all possible ways.
- Related to Within Cluster Resampling of Hoffman Sen & Weinberg (2001) and Multiple Outputation of Follmann Proschan & Leifer (2003).
- Talk has emphasized simple examples to ease understanding. But can be applied quite generally.
- Note: only works if Z does not change within a cluster.
- Provides for valid inference if cluster size is related to outcome.