Order restricted hypothesis testing for three-arm clinical trials

Muna Arephin

30th Annual Conference of the International Society for Clinical Biostatistics, Prague

23 - 27 August 2009

CANCER RESEARCH UK

Barts and The London
Queen Mary's School of Medicine and Dentistry

Acknowledgements

Professor Peter Sasieni

Professor Jack Cuzick

Outline

- Clinical trials
- Model and hypotheses
- Test statistics
- Test performances
- Summary

Clinical trials

One or more active arms against a control arm (placebo or a standard treatment).

Clinical trials

One or more active arms against a control arm (placebo or a standard treatment).

Well-developed methodologies are available for comparing one active vs control.

Clinical trials

One or more active arms against a control arm (placebo or a standard treatment).

Well-developed methodologies are available for comparing one active vs control.

Increasing demand for multi-arm trials, due to cost effectiveness (and sometimes benefits of having fewer patients on control).

Clinical trials

One or more active arms against a control arm (placebo or a standard treatment).

Well-developed methodologies are available for comparing one active vs control.

Increasing demand for multi-arm trials, due to cost effectiveness (and sometimes benefits of having fewer patients on control).

Less methodology for comparing two or more active vs control.

Model and Notation

Let

$$Y_{ij} \sim N(\mu_i, 1),$$

where Y_{ij} is the scaled response from patient j on treatment $i = 0, 1, 2$, $j = 1, \ldots, n_i$.

Model and Notation

Let

$$Y_{ij} \sim N(\mu_i, 1),$$

where Y_{ij} is the scaled response from patient j on treatment $i = 0, 1, 2$, $j = 1, \ldots, n_i$.

Let

$$Z_i = \frac{\sqrt{N}(\bar{Y}_i - \bar{Y}_0)}{\sigma}, \; i = 1, 2,$$

where $N = n_0 + n_1 + n_2$, $\sigma^2 = \{1 - \delta\}/\{\delta(1 - 2\delta)\}$ and $\delta = n_1/N = n_2/N$.

Model and Notation

Then

$$\begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \Delta_1 \\ \Delta_2 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right),$$

where

$$\Delta_i = \frac{\sqrt{N}(\mu_i - \mu_0)}{\sigma}$$

and $\rho = Cov(Z_i, Z_j) = \delta/\{1-\delta\}$, $i \neq j$.

Hypotheses

For comparing three (or more) arms:

$H_0 : \Delta_1 \leq 0$ and $\Delta_2 \leq 0$ vs $H_1 : \Delta_i > 0$ for at least one $i = 1, 2$.

Hypotheses

For comparing three (or more) arms:

$H_0 : \Delta_1 \leq 0$ and $\Delta_2 \leq 0$ vs $H_1 : \Delta_i > 0$ for at least one $i = 1, 2$.

Note: H_0 is a composite hypothesis.

Hypotheses

For comparing three (or more) arms:

$H_0 : \Delta_1 \leq 0$ and $\Delta_2 \leq 0$ vs $H_1 : \Delta_i > 0$ for at least one $i = 1, 2$.

Note: H_0 is a composite hypothesis.

Whereas with two arms, a Z-test is uniformly most powerful (UMP), with three or more arms no UMP test is known.

Test statistics

We consider the following test statistics, where $A^+ = \max(0, A)$ and $X_1 = \max(Z_1, Z_2)$ and $X_2 = \min(Z_1, Z_2)$:

1. $T_\infty = X_1^+$;

Test statistics

We consider the following test statistics, where $A^+ = \max(0, A)$ and $X_1 = \max(Z_1, Z_2)$ and $X_2 = \min(Z_1, Z_2)$:

1. $T_\infty = X_1^+$;
2. $S_1^0 = Z_1 + Z_2 = X_1 + X_2$;

Test statistics

We consider the following test statistics, where $A^+ = \max(0, A)$ and $X_1 = \max(Z_1, Z_2)$ and $X_2 = \min(Z_1, Z_2)$:

1. $T_\infty = X_1^+$;
2. $S_1^0 = Z_1 + Z_2 = X_1 + X_2$;
3. $T_1 = X_1^+ + \frac{(X_2 - \rho X_1)^+}{\sqrt{1-\rho^2}}$;

Test statistics

We consider the following test statistics, where $A^+ = \max(0, A)$ and $X_1 = \max(Z_1, Z_2)$ and $X_2 = \min(Z_1, Z_2)$:

1. $T_\infty = X_1^+$;
2. $S_1^0 = Z_1 + Z_2 = X_1 + X_2$;
3. $T_1 = X_1^+ + \frac{(X_2 - \rho X_1)^+}{\sqrt{1-\rho^2}}$;
4. $T_2 = \sqrt{X_1^{+2} + \frac{(X_2 - \rho X_1)^{+2}}{1-\rho^2}}$.

Likelihood Ratio Test

T_2^2 is the likelihood ratio test statistic given by

$$T_2^2 = 2 \ln \left\{ \frac{L(\mathbf{y}; \hat{\boldsymbol{\mu}})}{L(\mathbf{y}; \tilde{\boldsymbol{\mu}})} \right\},$$

where $\hat{\boldsymbol{\mu}}$ is the unrestricted maximum likelihood estimator of $\boldsymbol{\mu}$ under $H_0 \cup H_1$;
and $\tilde{\boldsymbol{\mu}}$ is the restricted maximum likelihood estimator of $\boldsymbol{\mu}$ under H_0.

Likelihood Ratio Test

T_2^2 is the likelihood ratio test statistic given by

$$T_2^2 = 2 \ln \left\{ \frac{L(\mathbf{y}; \hat{\boldsymbol{\mu}})}{L(\mathbf{y}; \tilde{\boldsymbol{\mu}})} \right\},$$

where $\hat{\boldsymbol{\mu}}$ is the unrestricted maximum likelihood estimator of $\boldsymbol{\mu}$ under $H_0 \cup H_1$;
and $\tilde{\boldsymbol{\mu}}$ is the restricted maximum likelihood estimator of $\boldsymbol{\mu}$ under H_0.

Restricted model under H_0, $\mu_i \leq \mu_0, \; i = 1, 2$, defines a simple tree order.

Pool Adjacent Violators Algorithm (PAVA)

Assume $\hat{\mu}_1 > \hat{\mu}_2$ (other order follows by symmetry).

Pool Adjacent Violators Algorithm (PAVA)

Assume $\hat{\mu}_1 > \hat{\mu}_2$ (other order follows by symmetry).

Start with $\tilde{\mu}_i = \hat{\mu}_i, \ \forall i = 0, 1, 2 \ \Rightarrow T_2 = 0$.

Pool Adjacent Violators Algorithm (PAVA)

Assume $\hat{\mu}_1 > \hat{\mu}_2$ (other order follows by symmetry).

Start with $\tilde{\mu}_i = \hat{\mu}_i, \ \forall i = 0, 1, 2 \ \Rightarrow T_2 = 0$.

If $\hat{\mu}_1 > \hat{\mu}_0$,

- let $\tilde{\mu}_0 = \tilde{\mu}_1 = \frac{n_0 \hat{\mu}_0 + n_1 \hat{\mu}_1}{n_0 + n_1} \ \Rightarrow T_2 = X_1$.

Pool Adjacent Violators Algorithm (PAVA)

Assume $\hat{\mu}_1 > \hat{\mu}_2$ (other order follows by symmetry).

Start with $\tilde{\mu}_i = \hat{\mu}_i, \ \forall i = 0, 1, 2 \ \Rightarrow T_2 = 0$.

If $\hat{\mu}_1 > \hat{\mu}_0$,

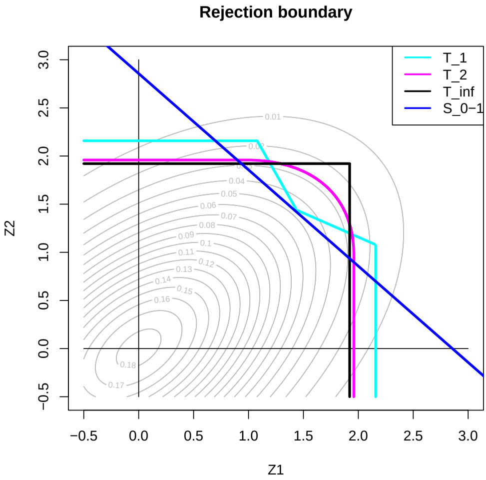
- let $\tilde{\mu}_0 = \tilde{\mu}_1 = \frac{n_0 \hat{\mu}_0 + n_1 \hat{\mu}_1}{n_0 + n_1} \ \Rightarrow T_2 = X_1$.
- If $\hat{\mu}_2 > \tilde{\mu}_0$,
  - let $\tilde{\mu}_i = \frac{n_0 \hat{\mu}_0 + n_1 \hat{\mu}_1 + n_2 \hat{\mu}_2}{n_0 + n_1 + n_2}, \ \forall i = 0, 1, 2 \ \Rightarrow T_2 = \sqrt{X_1^2 + \frac{(X_2 - \rho X_1)^2}{1 - \rho^2}}$.

Null distributions

The null distributions are approximated by Monte Carlo simulation of 1 million draws from the joint null distribution of Z_1 and Z_2.

Table: Critical values for rejecting H_0.

Test	5%	2.5%	1%
T_1	2.1537	2.5289	2.9839
T_2	1.9545	2.2579	2.6120
T_∞	1.9163	2.2121	2.5577
S_1^0	2.8488	3.3948	4.0291


Rejection boundary
T_1
T_2
T_inf
S_0−1
Z1
Z2
−0.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0.01
0.04
0.05
0.06
0.07
0.08
0.09
0.1
0.11
0.12
0.13
0.14
0.15
0.16
0.17
0.18

Test performances

We compare the different tests in terms of power and the expected loss by adding various values of Δ_1 and Δ_2 to 100,000 simulations from the joint null distribution of Z_1 and Z_2.

Test performances

We compare the different tests in terms of power and the expected loss by adding various values of Δ_1 and Δ_2 to 100,000 simulations from the joint null distribution of Z_1 and Z_2.

Power = the proportion of simulations which correctly rejected H_0.

Test performances

We compare the different tests in terms of power and the expected loss by adding various values of Δ_1 and Δ_2 to 100,000 simulations from the joint null distribution of Z_1 and Z_2.

Power $=$ the proportion of simulations which correctly rejected H_0.

Incorrect selection:
correctly reject H_0 but select the wrong experimental arm.

Test performances

We compare the different tests in terms of power and the expected loss by adding various values of Δ_1 and Δ_2 to 100,000 simulations from the joint null distribution of Z_1 and Z_2.

Power = the proportion of simulations which correctly rejected H_0.

Incorrect selection:
correctly reject H_0 but select the wrong experimental arm.

Expected Loss = the mean Loss over simulations, where

$$\text{Loss} = \begin{cases} 0 & \text{if we select treatment 1;} \\ \Delta_1 - \Delta_2 & \text{if we select treatment 2;} \\ \Delta_1 & \text{if we fail to reject } H_0. \end{cases}$$

Test performances

We compare the different tests in terms of power and the expected loss by adding various values of Δ_1 and Δ_2 to 100,000 simulations from the joint null distribution of Z_1 and Z_2.

Power $=$ the proportion of simulations which correctly rejected H_0.

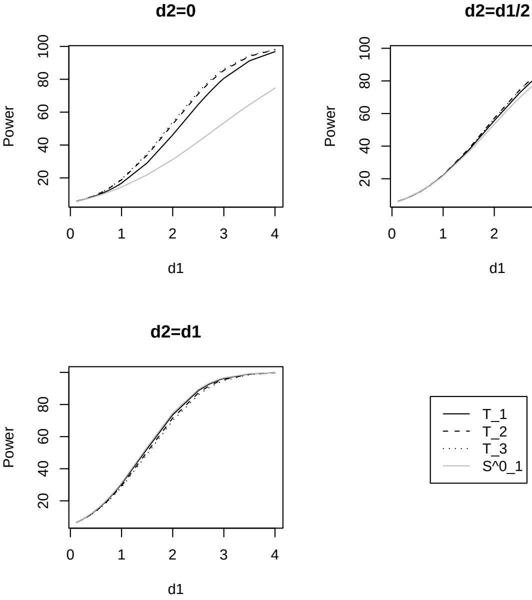
Incorrect selection:
correctly reject H_0 but select the wrong experimental arm.

Expected Loss $=$ the mean Loss over simulations, where

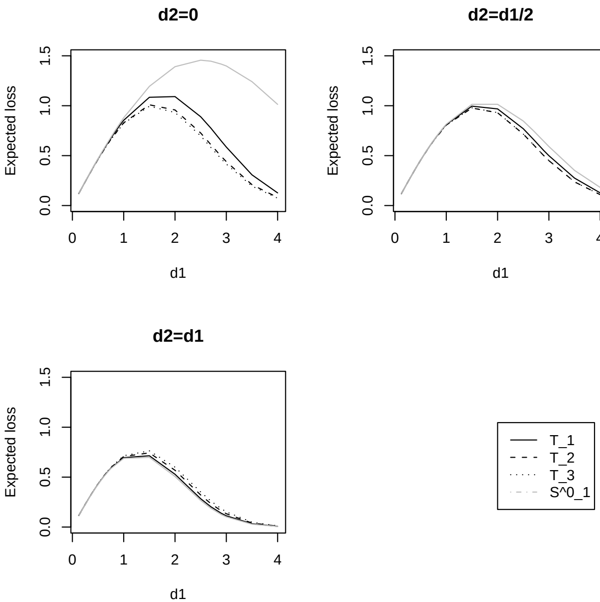
$$\text{Loss} = \begin{cases} 0 & \text{if we select treatment 1;} \\ \Delta_1 - \Delta_2 & \text{if we select treatment 2;} \\ \Delta_1 & \text{if we fail to reject } H_0. \end{cases}$$

We consider correlation $\rho = 0.5$, i.e. assuming equal numbers of patients in each arm.

Power at 5%



Expected loss at 5%



Summary

- No test is uniformly better than the others

Summary

- No test is uniformly better than the others
- T_∞ is best when only one treatment is better than control

Summary

- No test is uniformly better than the others
- T_∞ is best when only one treatment is better than control
- T_1 is better when treatments are similar

Summary

- No test is uniformly better than the others
- T_∞ is best when only one treatment is better than control
- T_1 is better when treatments are similar
- T_2 is more consistent with the best one in each of the above cases and is best in the middle

Summary

- No test is uniformly better than the others
- T_∞ is best when only one treatment is better than control
- T_1 is better when treatments are similar
- T_2 is more consistent with the best one in each of the above cases and is best in the middle
- Overall, there is not much difference between T_2 and T_∞

Summary

- No test is uniformly better than the others
- T_∞ is best when only one treatment is better than control
- T_1 is better when treatments are similar
- T_2 is more consistent with the best one in each of the above cases and is best in the middle
- Overall, there is not much difference between T_2 and T_∞
- T_2 has rarely (if ever) been used, but is at least as good as T_∞

Summary

- No test is uniformly better than the others
- T_∞ is best when only one treatment is better than control
- T_1 is better when treatments are similar
- T_2 is more consistent with the best one in each of the above cases and is best in the middle
- Overall, there is not much difference between T_2 and T_∞
- T_2 has rarely (if ever) been used, but is at least as good as T_∞

Summary

- No test is uniformly better than the others
- T_∞ is best when only one treatment is better than control
- T_1 is better when treatments are similar
- T_2 is more consistent with the best one in each of the above cases and is best in the middle
- Overall, there is not much difference between T_2 and T_∞
- T_2 has rarely (if ever) been used, but is at least as good as T_∞

The results can be extended to more arms.

References

Barlow, Bartholomew, Bremner and Brunk (1972) *Statistical Inference Under Order Restrictions*. Wiley.

Robertson, Wright, Dykstra (1988) *Order Restricted Statistical Inference*. Wiley.