Data combination in seamless Phase II/III designs

Christopher Jennison

Department of Mathematical Sciences,

University of Bath

http://people.bath.ac.uk/mascj

and **Lisa Hampson**

Department of Mathematics and Statistics,

Lancaster University

Takeda, London

January 2015

Outline of talk

1. Seamless Phase II/III clinical trials
2. Example: Formulation of the selection and testing problem
3. Six testing procedures
4. Power curves for the six tests
5. Understanding tests' performance via an "optimal" test
6. Assessments under other configurations of the treatment effects
7. The value of Phase II data in the final analysis
8. Optimal division of sample size between Phases II and III
9. Comments and conclusions

Reference: Hampson and Jennison (2015) Optimizing the data combination rule for seamless phase II/III clinical trials, *Statistics in Medicine*, **37**, 39–58.

1. Phase II and III clinical trials

Phase IIa

A “proof of concept” study intended to show evidence of drug activity at some dose.

Phase IIb

A trial to compare several doses or other variants of a treatment against control, with the aim of selecting a treatment for testing in a “confirmatory” Phase III trial.

Phase III

A trial intended to demonstrate the treatment selected in Phase II is superior to the control.

Seamless Phase II/III trial

A “seamless” design combines Phase IIb and Phase III, selecting a treatment and continuing to test this against control in a single study.

Seamless Phase II/III trials

Joint planning

Combining Phases II and III promises a more rapid final decision since the “white space” between phases is eliminated.

This requires substantial preparation to anticipate all the eventualities at the end of Phase II and, in each case, document how the study should continue.

Combining data

We shall discuss the issue of combining information in Phase II and Phase III data.

This option is of interest when the patient population and primary endpoint remain the same over both phases.

But, the final hypothesis test will need proper consideration of the selection process in Phase II, which generated the hypothesis now being tested.

2. Example

Jennison & Turnbull (*J. Biopharm. Statistics*, 2007, Sec. 5) consider:

Phase II

Four treatments and a control are compared, with $m_1 = 100$ observations on each.

Estimated treatment effects are $\hat{\theta}_{1,i},\ i = 1, \ldots, 4$.

The treatment i^* with highest $\hat{\theta}_{1,i}$ is selected for Phase III.

Phase III

Treatment i^* is compared against control, with $m_2 = 500$ observations on each.

Estimated treatment effect is $\hat{\theta}_{2,i^*}$.

Conclusion

A final decision is made, based on $\hat{\theta}_{1,1}, \ldots, \hat{\theta}_{1,4}$ and $\hat{\theta}_{2,i^*}$.

Example: Requirements

There are four null hypotheses, H_i: $\theta_i \leq 0$, $i = 1, \ldots, 4$.

If dose i^* is selected for Phase III, we focus on testing H_{i^*}: $\theta_{i^*} \leq 0$.

Family-wise error

We wish to control **family-wise error**, so, for all vectors $\theta = (\theta_1, \ldots, \theta_4)$,

$$Pr\{\text{Reject } \textit{any} \text{ true } H_i\} \leq \alpha.$$

Then, the probability of falsely claiming significance for the selected i^* is at most α.

Power

When some of the θ_i are greater than zero, we wish to have a high probability of selecting an effective treatment and rejecting the associated null hypothesis.

Formally, define power as

$$Pr\{\text{Select the treatment } j \text{ with maximum } \theta_i \text{ and reject } H_j\text{: } \theta_j \leq 0\}$$

(other definitions of power are possible).

Example: Further details

Phase II

Responses follow distributions

$$N(\mu_i,\, \sigma^2), \text{ on treatments } 1 \text{ to } 4,$$

$$N(\mu_c,\, \sigma^2), \text{ on control,}$$

where $\sigma^2 = 25$.

Treatment effects are $\theta_i = \mu_i - \mu_c$.

The estimated effects are

$$\hat{\theta}_{1,i} = \hat{\mu}_{1,i} - \hat{\mu}_{1,c} \sim N(\theta_i, \tfrac{2\,\sigma^2}{m_1}),$$

with correlation 0.5 between each pair.

Early stopping

If $\hat{\theta}_{1,i^*} = \max_i(\hat{\theta}_{1,i}) < 0$, stop for futility, otherwise proceed to Phase III.

Example: Further details

Phase III

The estimated effect of treatment i^* in Phase III is

$$\hat{\theta}_{2,i^*} \;=\; \hat{\mu}_{2,i^*} - \hat{\mu}_{2,c} \;\sim\; N(\theta_{i^*}, \tfrac{2\,\sigma^2}{m_2}).$$

Question: How should one make the final decision based on

$$\hat{\theta}_{1,1}, \ldots, \hat{\theta}_{1,4} \quad \text{and} \quad \hat{\theta}_{2,i^*} \; ?$$

— How to adjust for multiplicity? How to weight data from the two phases?

Jennison & Turnbull (*J. Biopharm. Statistics*, 2007) consider several methods:

Conventional: Final decision is based on $\hat{\theta}_{2,i^*}$ alone.

Bauer and Köhne: Applying combination tests to data from the two phases.

TSE: A design of Thall, Simon & Ellenberg (*Biometrika*, 1988).

3.1. Conventional: Final test based on Phase III data alone

It is not necessary to adjust for data-driven selection of hypothesis H_{i^*} since this is to be tested using only the *new* data from Phase III.

So, we can define

$$Z = \frac{\hat{\theta}_{2,i^*}}{\sqrt{(2\,\sigma^2/m_2)}}$$

and reject H_{i^*}: $\theta_{i^*} \leq 0$ in favour of $\hat{\theta}_{i^*} > 0$ if

$$Z > \Phi^{-1}(1 - 0.025) = 1.96.$$

With stopping for futility if $\hat{\theta}_{1,i^*} < 0$, the overall type I error rate is only 0.020.

For type I error rate 0.025 when $m_1 = 100$ and $m_2 = 500$, reject H_{i^*} if

$$Z > 1.86.$$

3.2 Bauer & Köhne (*Biometrics*, 1994): Combination tests

Testing a null hypothesis H_0 against a one-sided alternative, we have p-values:

$$P_1 \text{ from stage 1 (Phase II), } P_2 \text{ from stage 2 (Phase III).}$$

Under H_0, P_1 and P_2 have independent $U(0, 1)$ distributions.

a) Inverse χ^2 test

$$-\ln(P_1\, P_2) \sim \frac{1}{2}\,\chi^2_4.$$

Hence, a size α test is obtained by rejecting H_0 if

$$-\ln(P_1\, P_2) > \frac{1}{2}\,\chi^2_{4,\,1-\alpha}.$$

This χ^2 test was originally proposed for combining results of several studies by R. A. Fisher (1932) *Statistical Methods for Research Workers*.

Bauer & Köhne: Combination tests

b) Inverse normal test

Let $Z_1 = \Phi^{-1}(1 - P_1)$ and $Z_2 = \Phi^{-1}(1 - P_2)$.

Pre-specify weights w_1 and w_2, where $w_1^2 + w_2^2 = 1$.

Then under H_0, $w_1 Z_1 + w_2 Z_2 \sim N(0, 1)$.

Hence, a size α test is obtained by rejecting H_0 if

$$w_1 Z_1 + w_2 Z_2 > \Phi^{-1}(1 - \alpha).$$

Multiple comparisons

Since the choice of the null hypothesis H_{i^*} is based on the data that generated P_1 and Z_1, we need to allow for multiple comparisons if we are to control the family-wise type I error rate.

Closed testing procedures (Marcus et al, *Biometrika*, 1976)

We have 4 null hypotheses, H_i: $\theta_i \leq 0$, for $i = 1, \ldots, 4$. For each subset I of $\{1, \ldots, 4\}$, we define the intersection hypothesis

$$H_I = \cap_{j \in I} H_j.$$

We construct a level α test of each intersection hypothesis H_I: this test rejects H_I with probability at most α whenever all hypotheses specified in H_I are true.

Closed testing procedure

The simple hypothesis H_i: $\theta_i \leq 0$ is rejected overall if, and only if, H_I is rejected for every set I containing index i.

Proof of strong control of family-wise error rate

Let $\tilde{I}$ be the set of indices of all true hypotheses H_i. For a family-wise error to be committed, $H_{\tilde{I}}$ must be rejected. Since $H_{\tilde{I}}$ is true, $Pr\{\text{Reject } H_{\tilde{I}}\} = \alpha$ and, thus, the probability of a family-wise error is no greater than α.

Closed testing procedures using combination tests

(Bretz et al. and Schmidli et al, *Biometrical Journal*, 2006)

To test $H_I = \cap_{j \in I} H_j$, combine p-values $P_{1,I}$ from stage 1 and $P_{2,I}$ from stage 2.

Stage 1 (Phase II)

If H_I is the intersection of m simple hypotheses, we have p-values $P_{1,j},\ j \in I,$ for the individual null hypotheses H_j.

These can be combined by:

The Bonferroni method, giving overall p-value $P_{1,I} = m \ \min_{j \in I}(P_{1,j})$.

Simes' method, which is similar but less conservative.

Dunnett's test for multiple comparisons with a control.

Stage 2 (Phase III)

We are only interested in hypotheses H_I where I contains i^* and we set the p-value for each of these to be $P_{2,I} = P_{2,i^*}$.

Testing an intersection hypothesis

Simes' method (Biometrika, 1986):

If H_I is the intersection of m simple hypotheses, denote the p-values for these hypotheses in increasing order by $P_{1,(k)}$, $k = 1, \ldots, m$. The P-value for H_I is

$$P_{1,I} = \min_{k=1,\ldots,m} (m\, P_{1,(k)}/k).$$

This method is valid — and slightly conservative — when the P_i are independent or positively dependent, as in a comparison of m treatments with a control.

If treatment i^* has the highest $\hat{\theta}_{1,i}$ and smallest P-value of all k treatments, we have $P_{1,(1)} = P_{1,i^*}$ in any set I containing i^*.

The term $m\, P_{1,(k)}/k$ with $k = 1$ becomes $m\, P_{1,i^*}$, the usual "Bonferroni adjusted" version of P_{1,i^*}.

Simes' method allows other low P-values to reduce the overall result: if a second treatment performs well, $P_{1,(2)}/2$ may be smaller than P_{1,i^*}, reducing $P_{1,I}$.

Testing an intersection hypothesis

Dunnett's method (J. American Statistical Assoc., 1955):

We are to test the intersection hypothesis H_I, where I has m elements.

Assuming normal responses with known variance, the P-value for H_I is based on the Z-statistics, Z_j, arising from tests of H_j, $j \in I$, which compare one treatment at a time against the control.

The least favourable case in H_I has $\theta_j = 0$ for all j. Then, each $Z_j \sim N(0,1)$ and Cov$(Z_j, Z_{j'}) = 0.5$ for $j \neq j'$.

Let $Z^* = \max_{j \in I} Z_j$ and let z^* be the value attained by Z^* . The P-value for testing H_I using Dunnett's test is

$$P_I \;=\; P\{\max_{j \in I} Z_j > z^*\}$$

under the above multivariate normal joint distribution for Z_j, $j \in I$.

This can be converted to an adjusted Z-value, $Z_I = \Phi^{-1}(1 - P_I)$.

Closed testing procedures using combination tests

At the end of Phase III, we reject H_{i^*} if the specified combination test rejects every H_I where I contains i^*.

a) Inverse χ^2 test

The inverse χ^2 test rejects H_I if

$$-\ln(P_{1,I}\, P_{2,I}) \;>\; \frac{1}{2}\,\chi^2_{4,\,1-0.025} \;=\; 5.572.$$

With $m_1 = 100$ and $m_2 = 500$, using Simes' test for each H_I leads to an overall type I error rate under $\theta = (0,\, 0,\, 0,\, 0)$ of 0.021, due to stopping for futility at Phase II and the conservatism of Simes' test.

Rejecting each H_I if $-\ln(P_{1,I}\, P_{2,I}) \;>\; 5.376$ gives an overall type I error rate of 0.025 under $\theta = (0,\, 0,\, 0,\, 0)$.

Similarly, an overall type I error rate of 0.025 is attained by applying a critical value of 5.529 when a Dunnett test is used for each H_I.

Closed testing procedures using combination tests

b) Inverse normal test

We take $w_1 = \sqrt{(m_1/(m_1+m_2)}$ and $w_2 = \sqrt{(m_2/(m_1+m_2)}$, where m_1 and m_2 are the Phase II and Phase III sample sizes per treatment.

The inverse normal test rejects H_I if

$$w_1 Z_{1,I} + w_2 Z_{2,I} > \Phi^{-1}(1-0.025) = 1.96.$$

With $m_1 = 100$ and $m_2 = 500$, using Simes' test for each H_I leads to an overall type I error rate under $\theta = (0, 0, 0, 0)$ of 0.020, due to stopping for futility at Phase II and the conservatism of Simes' test.

Rejecting each H_I if $w_1 Z_{1,I} + w_2 Z_{2,I} > 1.86$ gives an overall type I error rate of 0.025 under $\theta = (0, 0, 0, 0)$.

Similarly, an overall type I error rate of 0.025 is attained by applying a critical value of 1.95 when a Dunnett test is used for each H_I.

3.3 Thall, Simon & Ellenberg (*Biometrika*, 1988)

Define

$$Z_{1,i^*} = \frac{\hat{\theta}_{1,i^*}}{\sqrt{(2\,\sigma^2/m_1)}} \quad \text{and} \quad Z_{2,i^*} = \frac{\hat{\theta}_{2,i^*}}{\sqrt{(2\,\sigma^2/m_2)}}.$$

The hypothesis H_{i^*} is rejected if

$$w_1\, Z_{1,i^*} + w_2\, Z_{2,i^*} > 2.20.$$

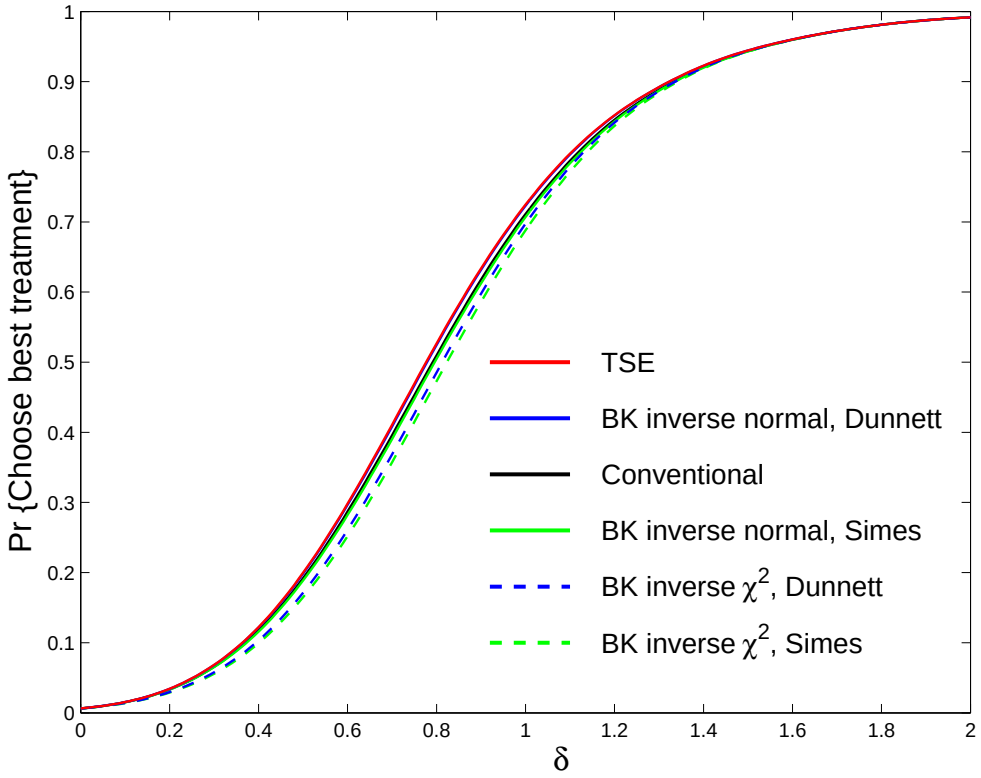
where $w_i = \sqrt{(m_i/(m_1+m_2)}$, $i = 1,\, 2,$ as before.

The critical value 2.20 is chosen to give overall type I error probability 0.025 if $\theta = (0,\, 0,\, 0,\, 0)$ and this guarantees a maximum family-wise error rate of 0.025.

It is easy to check the final decision is based on the difference in mean responses on treatment i^* and the control, pooled across Phases II and III — a very natural way in which to combine the relevant data.

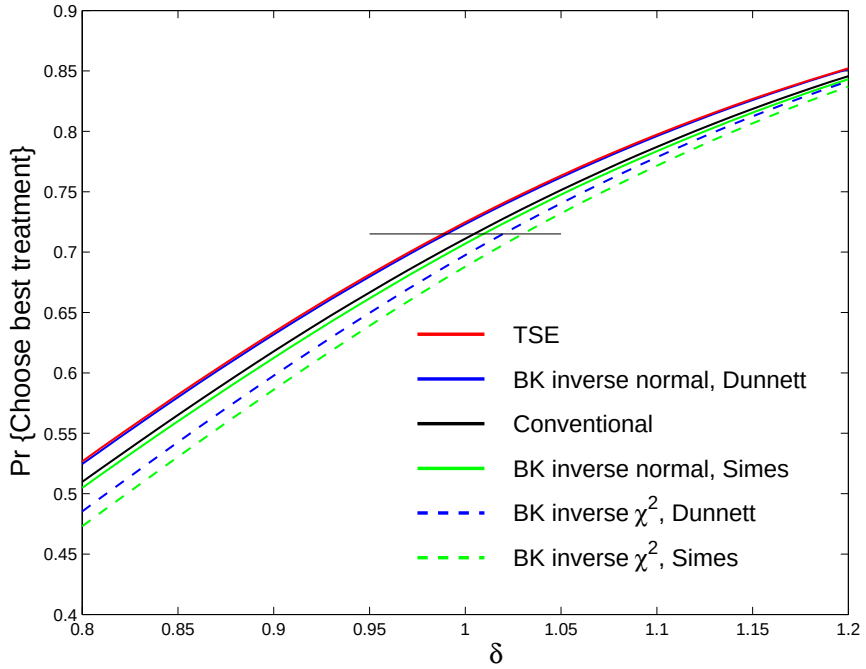
4. Power functions of the six selection and testing procedures

Power of six 2-stage selection/testing procedures when $\theta = (0,\ 0,\ 0,\ \delta)$,
i.e., three treatments are ineffective and the other has effect size δ:



NB The sample size distribution is the same for all six methods.

Power functions of the six selection and testing procedures



The Thall, Simon & Ellenberg design has slightly higher power than the inverse normal combination test using a Dunnett rule.

The conventional procedure, with no data combination, is less powerful than the TSE design — but superior to three versions of the BK combination test!

Differences in power equate to differences in sample size of around 4% to 8%.

5. Understanding the power curves

We shall seek an "optimal" procedure for this situation, examine its form, and compare with the six tests.

Requirements:

$Pr\{\text{Type I error}\} = 0.025$ when $\theta = (0, 0, 0, 0)$,

Maximum possible power when $\theta = (0, 0, 0, \delta)$ or a permutation of this.

To find the optimal procedure:

Consider a Bayes decision problem with:

Prior:

0.2 on $\theta = (0, 0, 0, 0)$,

0.2 on each permutation of $\theta = (0, 0, 0, \delta)$

Costs:

c_1 for rejecting any H_i when $\theta = (0, 0, 0, 0)$,

$-c_2$ for rejecting H_j when $\theta_j = \delta$.

An “optimal” procedure

The Bayes rule minimises

$$0.2 \times c_1 \times \{\text{Type I error probability under } \theta = (0,\, 0,\, 0,\, 0)\}$$

$$-4 \times 0.2 \times c_2 \times \{\text{Power at } \theta = (0,\, 0,\, 0,\, \delta)\}.$$

Hence, it maximises power at $\theta = (0,\, 0,\, 0,\, \delta)$ among all procedures with the same type I error probability under $\theta = (0,\, 0,\, 0,\, 0)$.

Choosing c_1 and c_2 so that the procedure has total type I error rate 0.025 under $\theta = (0,\, 0,\, 0,\, 0)$ will give the solution to our stated problem.

It may seem strange not to penalise, say, selecting treatment 2 and rejecting H_2 when $\theta = (0,\, 0,\, 0,\, \delta)$. However, family-wise error will be protected under this θ.

Most importantly, this formulation is what we need to solve the problem as posed.

An "optimal" procedure

Consider the final analysis after selecting treatment i^* and observing Phase III data.

Without loss of generality, suppose $i^* = 4$.

Possible costs are:

If $\theta = (0,\, 0,\, 0,\, 0)$	Reject H_{i^*} $\Rightarrow$ cost c_1
If $\theta = (0,\, 0,\, 0,\, \delta)$	Reject H_{i^*} $\Rightarrow$ cost $-c_2$
Other values of θ	No cost, whether H_{i^*} is rejected or accepted.

It follows that the Bayes optimal decision rule depends on the likelihood ratio of the Phase II and Phase III data under $\theta = (0,\, 0,\, 0,\, \delta)$ and $\theta = (0,\, 0,\, 0,\, 0)$.

To evaluate this likelihood ratio, recall

$$\hat{\theta}_{1,i} \sim N(\theta_i, \frac{2\,\sigma^2}{m_1}), \quad i = 1, \ldots, 4, \quad \hat{\theta}_{2,i^*} \sim N(\theta_{i^*}, \frac{2\,\sigma^2}{m_2})$$

and each pair $(\hat{\theta}_{1,i},\, \hat{\theta}_{1,j})$ has correlation 0.5.

An "optimal" procedure

Some algebra shows the optimal procedure rejects H_{i^*} if

$$\frac{8\,m_1}{m_1+m_2}\,\{\hat{\theta}_{1,i^*} \;-\; \sum_{j\neq i^*}\frac{1}{4}\,\hat{\theta}_{1,j}\} \;+\; \frac{5\,m_2}{m_1+m_2}\,\hat{\theta}_{2,i^*} \quad > \quad k$$

for some constant k.

The terms $\hat{\theta}_{1,j}$ have negative weights due to their positive correlations with $\hat{\theta}_{1,i^*}$.

Remember that for all the values of θ in our stated problem, three of the θ_j are equal to zero. Positive values of $\hat{\theta}_{1,j}$ for $j \neq i^*$ could be attributable to unusually negative responses on the common control arm, and this detracts from the significance of a high positive $\hat{\theta}_{1,i^*}$.

An "optimal" procedure

We have seen that the "optimal" procedure gives a negative weight to the Phase II estimates $\hat{\theta}_{1,j} \quad j \neq i^*$.

The Bauer & Köhne methods use closed testing procedures.

When these combine p-values by Simes rule to test intersection hypotheses, higher values of $\hat{\theta}_{1,j} \quad j \neq i^*$ contribute positively to rejection of H_{i^*}.

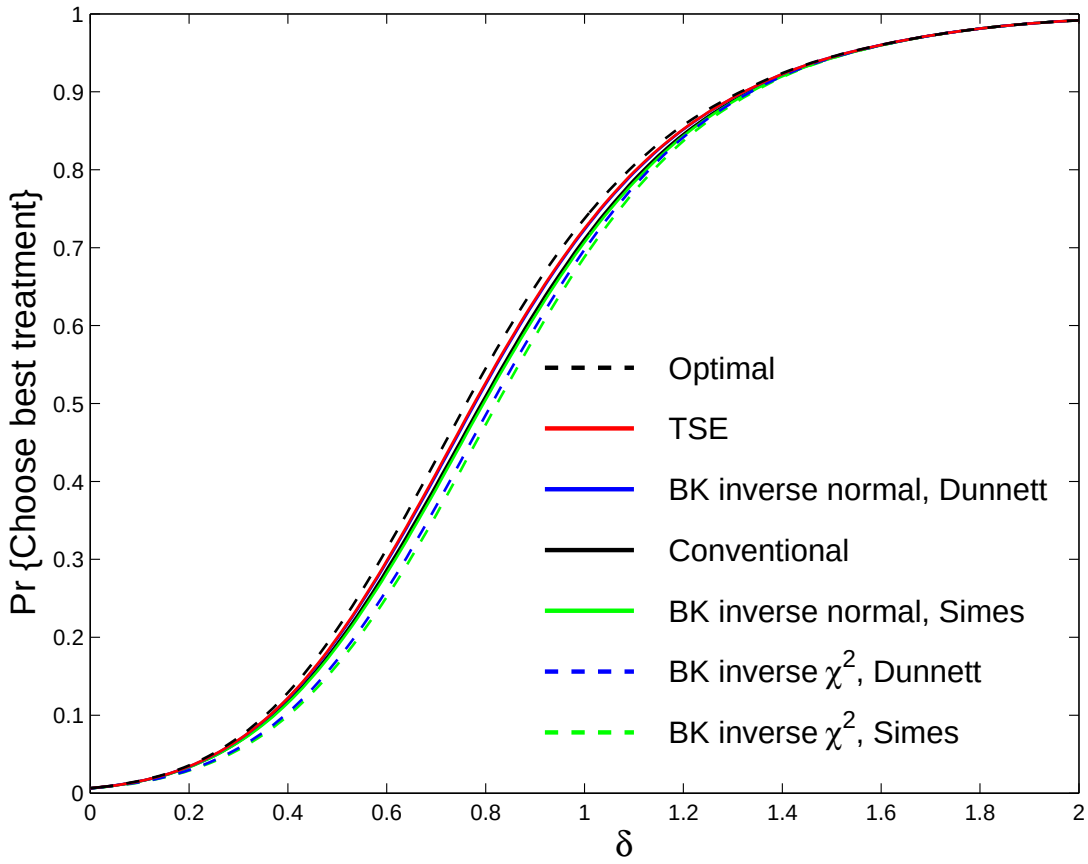
This is a natural way to "borrow strength" when one expects related treatments to have similar effects, but it is counter-productive when the other treatment effects are, in fact, all zero. Dunnett's test is much better suited to this situation.

The TSE procedure gives zero weight to the other $\hat{\theta}_{1,j}$s, and so is closer to the "optimal" decision rule.

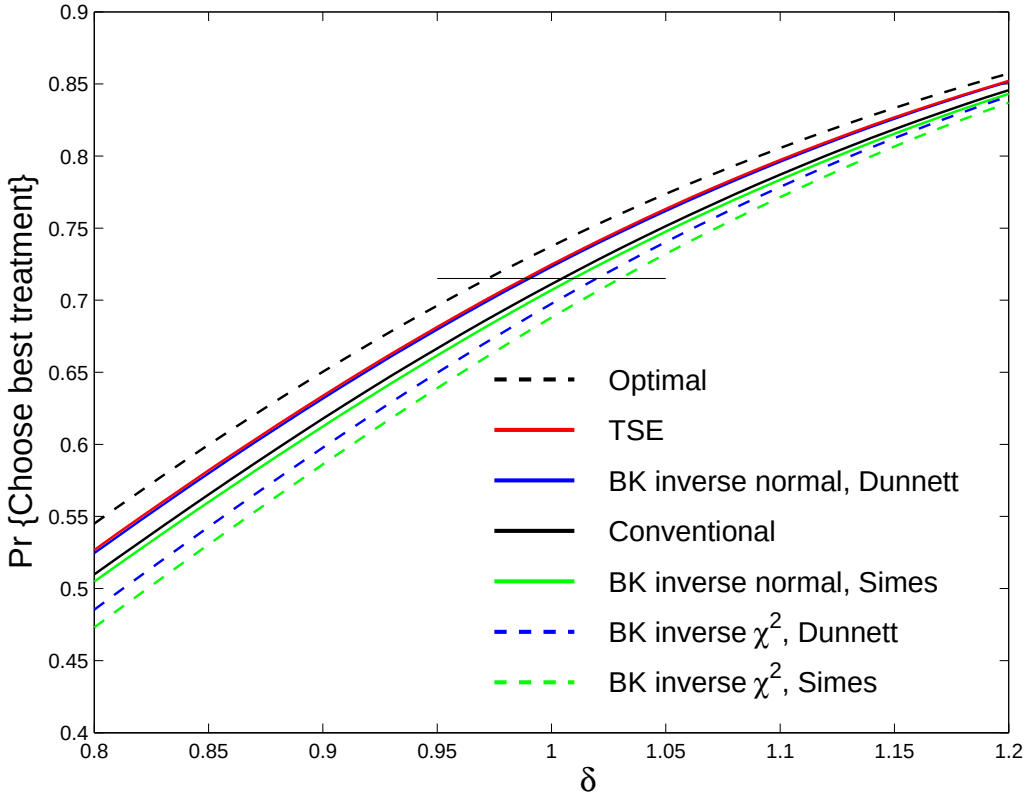
The conventional rule ignores all phase II estimates and at least it does not suffer from including terms $\hat{\theta}_{1,j}$ with the wrong sign.

Properties of six tests plus the “optimal”

Power of six 2-stage selection/testing procedures when $\theta = (0,\, 0,\, 0,\, \delta)$, plus the “optimal” procedure as a benchmark.



Properties of six tests plus the “optimal”



Differences in power equate to differences in sample size of around 5% to 10%.

But, there is a problem with the “optimal” procedure . . . .

The problem with the "optimal" procedure

This procedure rejects H_{i^*} if

$$\frac{8\,m_1}{m_1+m_2}\,\{\hat{\theta}_{1,i^*} \;-\; \sum_{j\neq i^*}\frac{1}{4}\,\hat{\theta}_{1,j}\} \;+\; \frac{5\,m_2}{m_1+m_2}\,\hat{\theta}_{2,i^*} \;\;>\;\; k.$$

Here, k is chosen so type I error probability is 0.025 when $\theta = (0,\, 0,\, 0,\, 0)$.

Negative weights for $\hat{\theta}_{1,j}$, $j \neq i^*$, imply high type I error probability for, say, $\theta = (-\lambda,\, -\lambda,\, -\lambda,\, 0)$ where $-\lambda$ is large and negative. So, the "optimal" test does not satisfy the family-wise error condition.

Question: What procedure is truly optimal, maximising power at $\theta = (0,\, 0,\, 0,\, \delta)$ and permutations of this while protecting the family-wise error rate?

Conjecture: Increasing the weights of $\hat{\theta}_{1,j}$, $j \neq i^*$, so they are non-negative (zero, in fact) and weighting $\hat{\theta}_{1,i^*}$ and $\hat{\theta}_{2,i^*}$ by the inverse of their variances gives the TSE rule — suggesting this should be close to optimal for this problem.

6. Performance of tests under other configurations of θ

Example 2:

Suppose θ is a permutation of $(\gamma\,\delta,\ \gamma\,\delta,\ \gamma\,\delta,\ \delta)$, where $0 \leq \gamma < 1$.

Aim:

High power to select the treatment i^* for which $\theta_{i^*} = \delta$ and then reject H_{i^*}.

Optimal rules

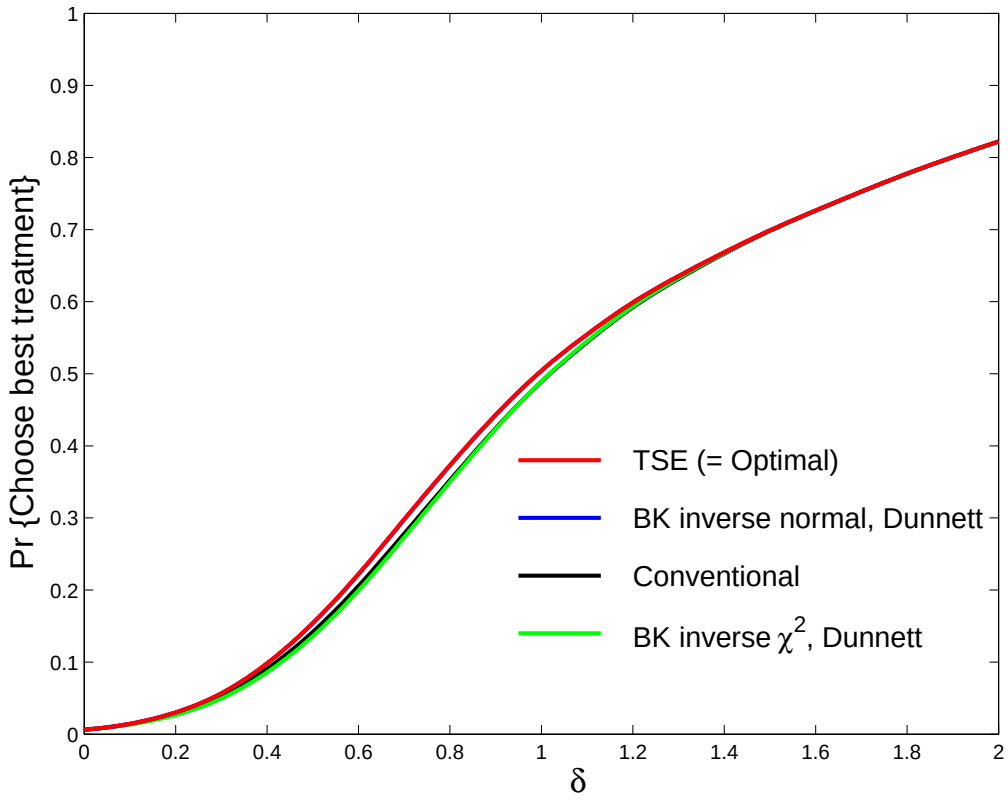
We can follow the same approach of solving a suitably formulated Bayes decision problem to derive an "optimal" rule.

This rule gives non-negative weights to all the $\hat{\theta}_{1,j}$s when $\gamma \geq 0.5$.

The optimal rule for $\gamma = 0.5$ is the TSE decision rule.

Example 2 with $\gamma = 0.5$

Power of four procedures when $\theta = (\gamma, \gamma, \gamma, 1)\,\delta$ with $\gamma = 0.5$.

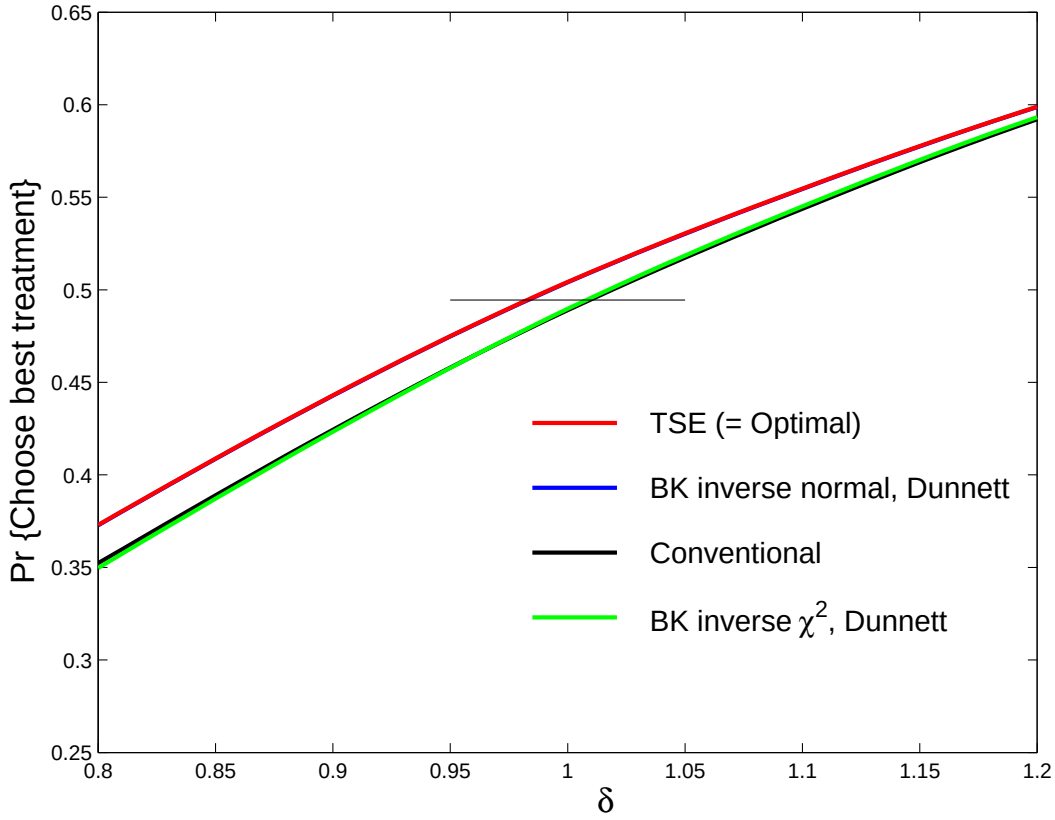


TSE is exactly optimal — but the BK-inverse normal-Dunnett test is very similar.

Dunnett BK tests are slightly superior to Simes versions (not displayed).

Example 2 with $\gamma = 0.5$

Power of four procedures when $\theta = (\gamma, \gamma, \gamma, 1)\,\delta$ with $\gamma = 0.5$.

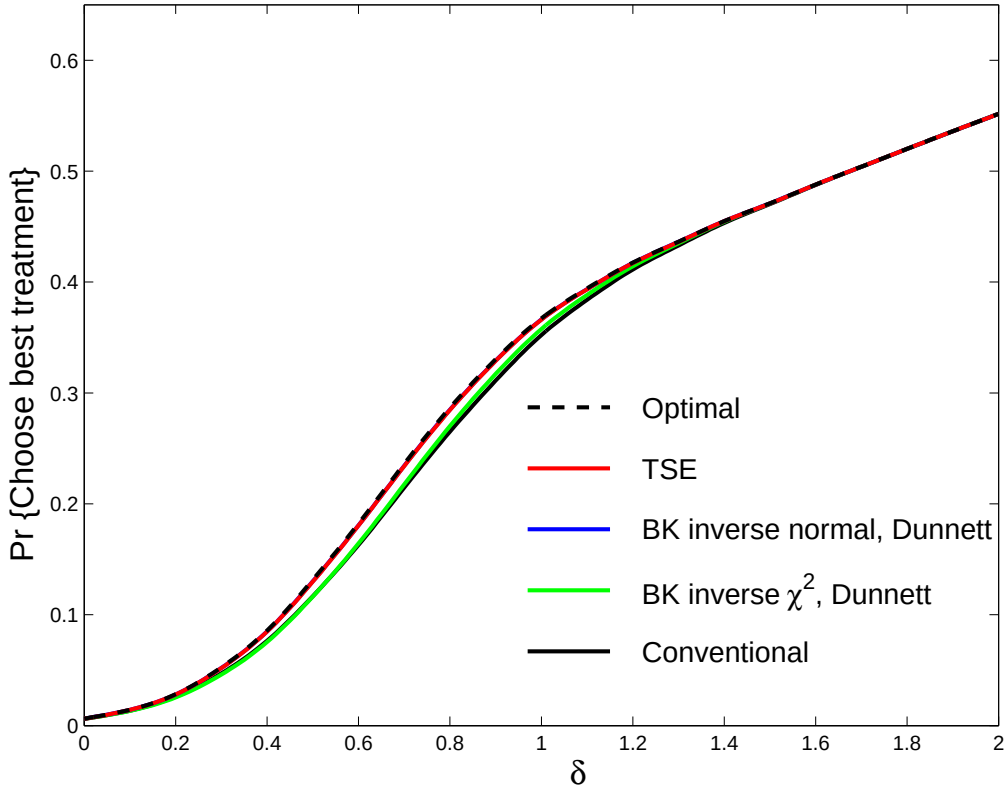


TSE is exactly optimal — but the BK-inverse normal-Dunnett test is very similar.

Dunnett BK tests are slightly superior to Simes versions (not displayed).

Example 2 with $\gamma = 0.75$

Power of four procedures when $\theta = (\gamma,\ \gamma,\ \gamma,\ 1)\,\delta$ with $\gamma = 0.75$, plus the "optimal" procedure for this case as a benchmark.

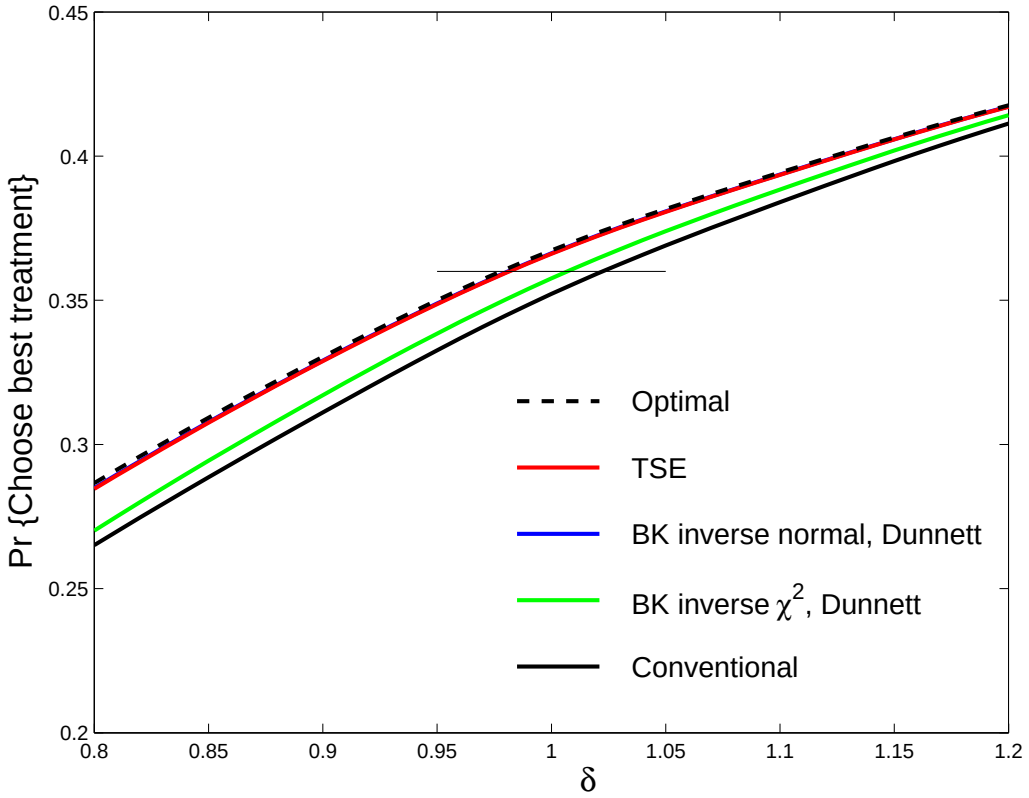


TSE and the BK-inverse normal-Dunnett test are close to optimal here.

Dunnett BK tests are slightly superior to Simes versions.

Example 2 with $\gamma = 0.75$

Power of four procedures when $\theta = (\gamma,\ \gamma,\ \gamma,\ 1)\,\delta$ with $\gamma = 0.75$, plus the "optimal" procedure for this case as a benchmark.



TSE and the BK-inverse normal-Dunnett test are close to optimal here.

Dunnett BK tests are slightly superior to Simes versions.

Performance of tests under other configurations of θ

Example 3:

Suppose θ is a permutation of $(\gamma_1, \gamma_2, \gamma_3, 1)\,\delta$, for $0 \leq \gamma_1 \leq \gamma_2 \leq \gamma_3 < 1$.

Aim:

High power to select the treatment i^* for which $\theta_{i^*} = \delta$ and then reject H_{i^*}.

Optimal rules

Again, we can follow a suitably formulated Bayes decision problem to derive an "optimal" rule, which now involves a combination of likelihood ratios.

For sufficiently high values of $\gamma_1, \ldots, \gamma_3$, the optimal rule gives non-negative weights to all the $\hat{\theta}_{1,j}$s and family-wise type I error is properly protected.

We find the TSE rule remains highly efficient relative to the optimal rule.

Performance of tests under other configurations of θ

Example 4:

Suppose θ is a permutation of $(\gamma_1, \gamma_2, \gamma_3, 1)\,\delta$, for $0 \leq \gamma_1 \leq \gamma_2 \leq \gamma_3 < 1$.

Suppose also it is thought likely that the θ_i increase monotonically with i — this could be a reasonable view when treatments represent increasing dose levels.

Aim:

High power to select the treatment i^* for which $\theta_{i^*} = \delta$ and then reject H_{i^*}.

Optimal rules

Here, it is of interest to find the "optimal" rule based on the assumption that the θ_i increase monotonically with i.

Once again, a suitably formulated Bayes decision problem leads to an "optimal" rule.

Performance of tests under other configurations of θ

Example 4 continued:

Treatment effects are a permutation of $\theta = (\gamma_1, \gamma_2, \gamma_3, 1)\,\delta$ *and the* θ_i *are expected to increase monotonically with* i*.*

Calculations show that the additional knowledge about the monotonicity of the θ_is provides only a very slight efficiency gain.

Thus, the TSE rule remains highly efficient relative to this specialised optimal rule.

This leads us to the conclusion that using a good dose-response model does not greatly improve the final decision to accept or reject H_{i^*}.

Bretz, Pinheiro & Branson (*Biometrics*, 2005) proposed the *MCP-Mod method* for adaptively choosing a model and defining the best contrast of estimated effects. Our results suggest this may not be necessary!

However, modelling may well be useful in adaptive allocation of treatments (doses) during the Phase II stage and data-based choice of Phase III sample size.

Relative efficiencies of decision rules

The relative efficiency of one procedure versus another can be defined in terms of the ratio of sample sizes needed for the procedures to have equal power.

Suppose Procedure A has power ϕ for given m_1 and m_2 and this is the highest power attained by any procedure.

Then Procedure A is 100% efficient.

Suppose Procedure B attains power ϕ if it is given sample sizes $\rho\, m_1$ and $\rho\, m_2$.

The we say the efficiency of Procedure B is

$$\frac{1}{\rho} \times 100\%.$$

Relative efficiencies of 6 rules: $m_1 = 100,\ m_2 = 500$

Decision rule	*Configuration of* θ				
	A	B	C	D	E
TSE	100	100	99	99	99
BK inverse normal, Dunnett	100	100	100	99	99
BK inverse χ^2, Dunnett	94	95	95	94	94
BK inverse normal, Simes	96	99	99	99	99
BK inverse χ^2, Simes	92	94	94	94	94
Phase III data only	97	95	92	91	90

A: $\theta = (0,\ 0,\ 0,\ \delta)$

B: $\theta = (\gamma,\ \gamma,\ \gamma,\ 1)\,\delta,\ \gamma = 0.5$

C: $\theta = (\gamma,\ \gamma,\ \gamma,\ 1)\,\delta,\ \gamma = 0.75$

D: $\theta = (\gamma_1,\ \gamma_2,\ \gamma_3,\ 1)\,\delta,\ \gamma_1 = 0.75$

E: $\theta = (\gamma_1,\ \gamma_2,\ \gamma_3,\ 1)\,\delta,\ \gamma_1 = 0.75,\ \theta$ assumed monotone.

Relative efficiencies of 6 rules: $m_1 = 200,\ m_2 = 400$

Decision rule	*Configuration of* θ				
	A	B	C	D	E
TSE	100	100	99	98	97
BK inverse normal, Dunnett	99	100	99	98	98
BK inverse χ^2, Dunnett	95	97	96	95	95
BK inverse normal, Simes	92	98	99	99	98
BK inverse χ^2, Simes	93	96	96	95	95
Phase III data only	88	87	84	81	81

A: $\theta = (0,\ 0,\ 0,\ \delta)$

B: $\theta = (\gamma,\ \gamma,\ \gamma,\ 1)\,\delta,\ \gamma = 0.5$

C: $\theta = (\gamma,\ \gamma,\ \gamma,\ 1)\,\delta,\ \gamma = 0.75$

D: $\theta = (\gamma_1,\ \gamma_2,\ \gamma_3,\ 1)\,\delta,\ \gamma_1 = 0.75$

E: $\theta = (\gamma_1,\ \gamma_2,\ \gamma_3,\ 1)\,\delta,\ \gamma_1 = 0.75,\ \theta$ assumed monotone.

7. Value of Phase II data in the final analysis

Hampson & Jennison consider how many additional Phase III subjects are needed to achieve the same increase in power gained by using Phase II data in the final analysis.

If this number of additional Phase III subjects is $2m_2^*$, the Phase II subjects on selected treatment i^* and control can be said to provide r% of their "face value", where

$$r \;=\; 100\,\frac{2\,m_2^*}{2\,m_1}.$$

In their example, using the TSE rule, this value ranges between 20% and 100%.

However, values in the range 50% to 70% are typical for the most commonly occurring vectors of treatment effects.

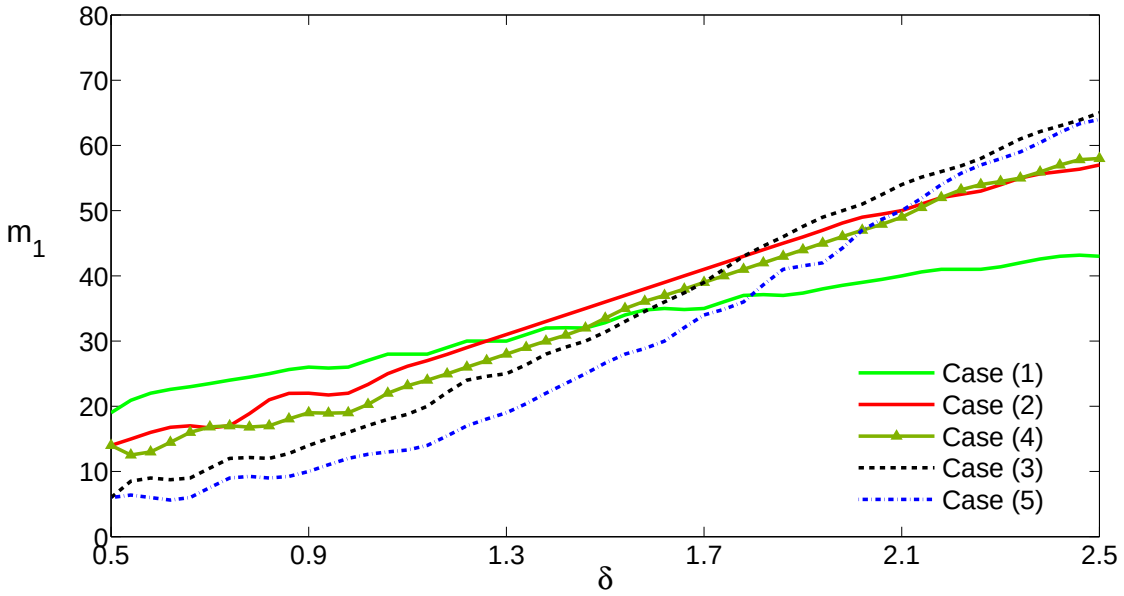
8. Optimal division of sample size between Phases II and III

Hampson & Jennison follow TSE's example in discussing the choice of group sizes m_1 and m_2 when the total sample size is fixed at $(K+1)\,m_1 + 2\,m_2 \;=\; N$.

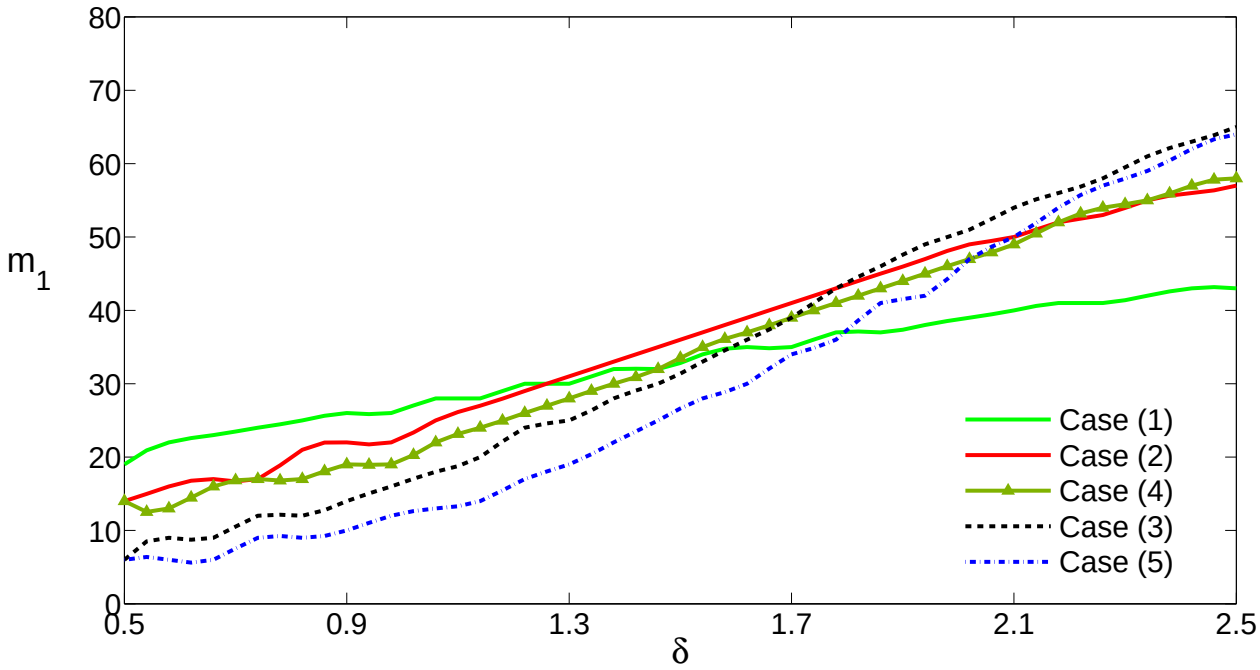
Using the TSE rule, H & J find values of m_1 and m_2 which lead to the maximum possible power for fixed N and a given vector of treatment effects, θ.

Their example has $K = 5$ treatments plus control, $\sigma^2 = 25$, and $N = 448$.

Optimal values of m_1 for five configurations of θ are plotted below.



Optimal division of sample size between Phases II and III



Hampson & Jennison conclude that no single choice of m_1 is ideal for all scenarios.

Rather, investigators should consider the most likely scenarios for their trial and choose group sizes to give high average power across these cases.

Alternative definitions of “power” should also be considered if two or more treatments may have high effect sizes.

9. Conclusions

Efficiency gains can be made by combining Phase II and Phase III data (more so for larger Phase II sample sizes).

Rules (i) *TSE* and (ii) *BK inverse normal combination rule with Dunnett's test* have robust efficiency over a variety of θ configurations, numbers of treatments, and sample sizes.

The TSE can be represented as a closed testing procedure (Jennison & Turnbull, *Biometrical Journal*, 2006). Hence, both this and the Dunnett procedure can still be applied when a treatment other than that with the highest $\hat{\theta}_{1,i}$ is selected for Phase III (e.g., for safety considerations).

We adjusted critical values for Bauer & Köhne procedures using Simes' test to eliminate conservatism under $\theta = (0,\, 0,\, 0,\, 0)$. This could conceivably increase the family-wise error rate for other θ configurations — but this only strengthens our recommendation of the TSE and Dunnett rules.

Conclusions

Multiple comparison procedures

The notion of “borrowing strength” implicit in Simes’ rule may not always be desirable for multiple comparisons with a control.

In contrast, the Dunnett procedure, which is designed for multiple comparisons with a control, performs well in this problem.

Combining data from Phases II and III

The TSE and BK-inverse normal-Dunnett procedures protect family-wise type I error and can be employed flexibly.

These procedures are robustly efficient and even have good efficiency when compared to model based methods which assume the correct model.

These procedures only use observations on the selected treatment and control in the final decision — with adjustment for data-dependent treatment selection.

Conclusions

The “value” of the Phase II data making in the final decision can be assessed by comparing efficiency against that of the decision rule based on Phase III data only.

Our results show the Phase II data on treatment i^* and the control are, typically, worth 50% to 70% of their face value: for example, if Phase II has 100 observations per treatment and control, these improve power by the same amount as an extra 50 to 70 observations on treatment i^* and control in Phase III.

The requirement by regulators to treat the combined study as a single trial means that issues usually addressed in the gap between Phases II and III must be anticipated and rules for how to proceed set up in the overall protocol.

Although Phase II data are of reduced value and their use in the Phase III analysis has an administrative cost, this practice may still be desirable when observations are at a premium, e.g., in a rare illness with slow patient recruitment.

Appendix 1: Likelihood ratios for a linear model

In comparing k treatments with a control, suppose $\theta_1 = \gamma_1\,\delta,\ \ldots\ ,\ \theta_k = \gamma_k\,\delta,$ so

$$\theta \;=\; \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_k \end{pmatrix} \;=\; \begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_k \end{pmatrix} \delta = X\delta$$

for a scalar parameter δ.

In Phase II, we have $\hat{\theta} \;\sim\; N(\theta,\,\Sigma) \;=\; N(X\delta,\,\Sigma)$, where

$$\Sigma \;=\; \begin{pmatrix} 2 & 1 & \ldots & 1 \\ 1 & 2 & \ldots & 1 \\ \vdots & \vdots & & \vdots \\ 1 & 1 & \ldots & 2 \end{pmatrix} \frac{\sigma^2}{m_1}.$$

Likelihood ratios for a linear model

The probability density of $\hat{\theta}$ is

$$f_\delta(\hat{\theta}) \;=\; (2\pi)^{-k/2}\,(\det\Sigma)^{-1/2}\;\exp\{-\frac{1}{2}(\hat{\theta}-X\delta)^T\Sigma^{-1}(\hat{\theta}-X\delta)\}.$$

The likelihood ratio for $\delta=\delta^*$ vs $\delta=0$ is

$$\frac{f_{\delta^*}}{f_0} \;=\; \exp(\delta^* X^T\Sigma^{-1}\hat{\theta} \;+ \textit{constant}\,)$$

and the log likelihood ratio is

$$\delta^* X^T\Sigma^{-1}\hat{\theta} \;+ \textit{constant}. \qquad (1)$$

Now, the maximum likelihood estimate of δ for this normal linear model is

$$\hat{\delta} \;=\; (X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1}\hat{\theta} \;\sim\; N(\delta,\,(X^T\Sigma^{-1}X)^{-1}).$$

Likelihood ratios for a linear model

So, the log likelihood ratio (1) is a constant plus

$$\delta^*(X^T\Sigma^{-1}X)\hat{\delta} \;=\; \delta^*\frac{\hat{\delta}}{\mathrm{Var}(\hat{\delta})}.$$

Denote the Phase II estimate of δ by $\hat{\delta}_1$ so the log likelihood ratio from Phase II data is

$$\delta^*\frac{\hat{\delta}_1}{\mathrm{Var}(\hat{\delta}_1)} \;+\; \text{constant.}$$

Similarly (with a simpler model as only $\hat{\theta}_{i^*}$ is observed), the log likelihood ratio from Phase III data is

$$\delta^*\frac{\hat{\delta}_2}{\mathrm{Var}(\hat{\delta}_2)} \;+\; \text{constant.}$$

Likelihood ratios for a linear model

Combining across Phase II and Phase III, the log likelihood ratio for $\delta = \delta^*$ vs $\delta = 0$ is a constant plus

$$\delta^* \{ \frac{\hat{\delta}_1}{\text{Var}(\hat{\delta}_1)} + \frac{\hat{\delta}_2}{\text{Var}(\hat{\delta}_2)} \}.$$

Thus, we reject $\delta = 0$ in favour of $\delta > 0$ for high values of

$$\frac{\hat{\delta}_1}{\text{Var}(\hat{\delta}_1)} + \frac{\hat{\delta}_2}{\text{Var}(\hat{\delta}_2)},$$

the overall MLE of δ from the pooled Phase II and Phase III data.

Appendix 2: Formula for the likelihood ratio from Phase II

The log likelihood ratio is an increasing function of $X^T \Sigma^{-1} \hat{\theta}$, i.e., of

$$(\gamma_1, \ldots, \gamma_k) \, \frac{m_1}{\sigma^2 (k+1)} \begin{pmatrix} k & -1 & \ldots & -1 \\ -1 & k & \ldots & -1 \\ \vdots & \vdots & & \vdots \\ -1 & -1 & \ldots & k \end{pmatrix} \begin{pmatrix} \hat{\theta}_{1,1} \\ \hat{\theta}_{1,2} \\ \vdots \\ \hat{\theta}_{1,k} \end{pmatrix}$$

$$= \frac{m_1}{\sigma^2} \, (\gamma_1, \ldots, \gamma_k) \, (I_n - \frac{1}{k+1} \, 1_{n \times n}) \begin{pmatrix} \hat{\theta}_{1,1} \\ \vdots \\ \hat{\theta}_{1,k} \end{pmatrix}$$

$$= \frac{m_1}{\sigma^2} \, \{ \sum_{j=1}^{k} \gamma_j \, \hat{\theta}_{1,j} \, - \, \frac{1}{k+1} \sum_{j=1}^{k} \gamma_j \, \sum_{j+1}^{k} \hat{\theta}_{1,j} \, \}. \qquad (2)$$

Special cases of the likelihood ratio from Phase II

Case 1.

With $i^* = k$, consider testing $\theta = (0, \ldots, 0)$ vs $\theta = (0, \ldots, 0, \delta)$.

We take $(\gamma_1, \ldots, \gamma_k) = (0, \ldots, 0, 1)$, then (2) becomes

$$\frac{m_1}{\sigma^2} \left(\hat{\theta}_{1,i^*} - \frac{1}{k+1} \sum_{j=1}^{k} \hat{\theta}_{1,j} \right)$$

$$= \frac{m_1}{\sigma^2} \frac{k}{k+1} \left(\hat{\theta}_{1,i^*} - \frac{1}{k} \sum_{j \neq i^*} \hat{\theta}_{1,j} \right) .$$

This lies behind the term

$$\hat{\theta}_{1,i^*} - \frac{1}{4} \sum_{j \neq i^*} \hat{\theta}_{1,j}$$

in the formula on Slide 24 for the "optimal" test with $k = 4$.

Special cases of the likelihood ratio from Phase II

Case 2.

With $i^* = k$, consider testing $\theta = (0, \ldots, 0)$ vs $\theta = (\gamma\delta, \ldots, \gamma\delta, \delta)$.

We take $(\gamma_1, \ldots, \gamma_k) = (\gamma, \ldots, \gamma, 1)$, then (2) becomes

$$\frac{m_1}{\sigma^2} \{ \hat{\theta}_{1,i^*} + \sum_{j \neq i^*} \gamma \hat{\theta}_{1,j} - \frac{(k-1)\gamma + 1}{k+1} (\hat{\theta}_{1,i^*} + \sum_{j \neq i^*} \hat{\theta}_{1,j}) \}$$

$$= \frac{m_1}{\sigma^2(k+1)} \{ [k - (k-1)\gamma] \hat{\theta}_{1,i^*} + \sum_{j \neq i^*} (2\gamma - 1) \hat{\theta}_{1,j} \}.$$

The coefficient $(2\gamma - 1)$ of terms $\hat{\theta}_{1,j}$ for $j \neq i^*$ is negative when $\gamma < 0.5$ and positive when $\gamma > 0.5$.

The case $\gamma = 0.5$ gives a coefficient of zero — and the TSE decision rule.

Appendix 3: Alternative derivation of MLEs in Phase II

Our original model assumed independent treatment estimates

$$\hat{\mu}_{1,i} \sim N(\mu_i, \frac{\sigma^2}{m_1}), \quad i = 1, \ldots, k,$$

$$\hat{\mu}_{1,c} \sim N(\mu_c, \frac{\sigma^2}{m_1}).$$

We defined $\theta_i = \mu_i - \mu_c$, $i = 1, \ldots, k$, and worked with the estimates $\hat{\theta}_{1,i}$.

However, we can also work with the $\hat{\mu}_{1,i}$s and $\hat{\mu}_c$ directly.

Example 1

We can write the case $\theta = (0, \ldots, 0, \delta)$ as

$$\begin{aligned} \mu_1 = \ldots = \mu_{k-1} = \mu_c &= a, \\ \mu_k &= a + \delta. \end{aligned}$$

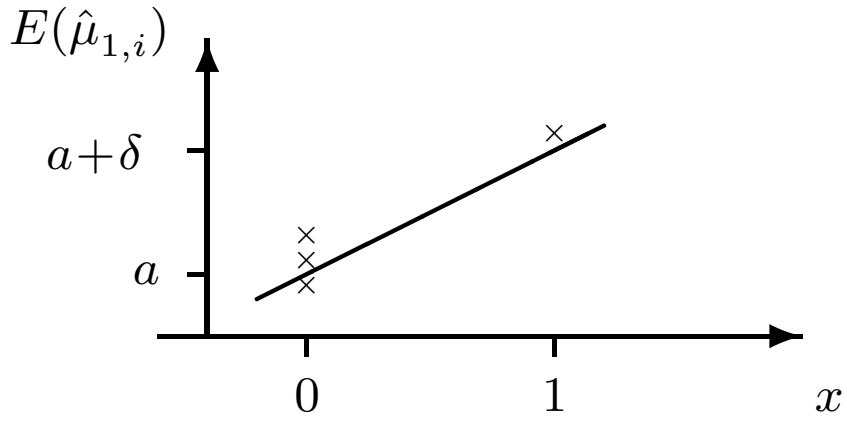
Alternative derivation of MLEs in Phase II

So, estimates follow a regression model $E(\hat{\mu}_{1,i}) = a + \delta\, x$ where

$\hat{\mu}_{1,1}, \ldots, \hat{\mu}_{1,k-1}$ and $\hat{\mu}_c$ have $x = 0$,

$\hat{\mu}_{1,k}$ has $x = 1$,

and we are interested in the slope δ.



Clearly,

$$\hat{\delta} = \hat{\mu}_{1,k} - \frac{1}{k}\,(\hat{\mu}_{1,1} + \ldots + \hat{\mu}_{1,k-1} + \hat{\mu}_{1,c}).$$

Alternative derivation of MLEs in Phase II

But, we can write

$$\begin{aligned}
\hat{\delta} &= \hat{\mu}_{1,k} - \frac{1}{k}\left(\hat{\mu}_{1,1} + \ldots + \hat{\mu}_{1,k-1} + \hat{\mu}_{1,c}\right) \\
&= (\hat{\mu}_{1,k} - \hat{\mu}_{1,c}) - \frac{1}{k}\sum_{j=1}^{k-1}(\hat{\mu}_{1,j} - \hat{\mu}_{1,c}) \\
&= \hat{\theta}_{1,k} - \frac{1}{k}\sum_{j=1}^{k-1}\hat{\theta}_{1,j},
\end{aligned}$$

in agreement with previous results.

Note: In the estimate of a slope, the sum of the weights of the $\hat{\mu}_{1,i}$s and $\hat{\mu}_c$ is zero (it is a contrast).

So, it is automatically the case that $\hat{\delta}$ can be expressed in terms of the $\hat{\theta}_{1,i}$s.

Alternative derivation of MLEs in Phase II

Example 2

We can write the case $\theta = (\gamma, \ldots, \gamma, 1)\,\delta$ as

$$\begin{aligned} \mu_c &= a, \\ \mu_1 = \ldots = \mu_{k-1} &= a + \gamma\,\delta, \\ \mu_k &= a + \delta. \end{aligned}$$

Now, the estimates follow a regression model $E(\hat{\mu}_{1,i}) = a + \delta\, x$ where

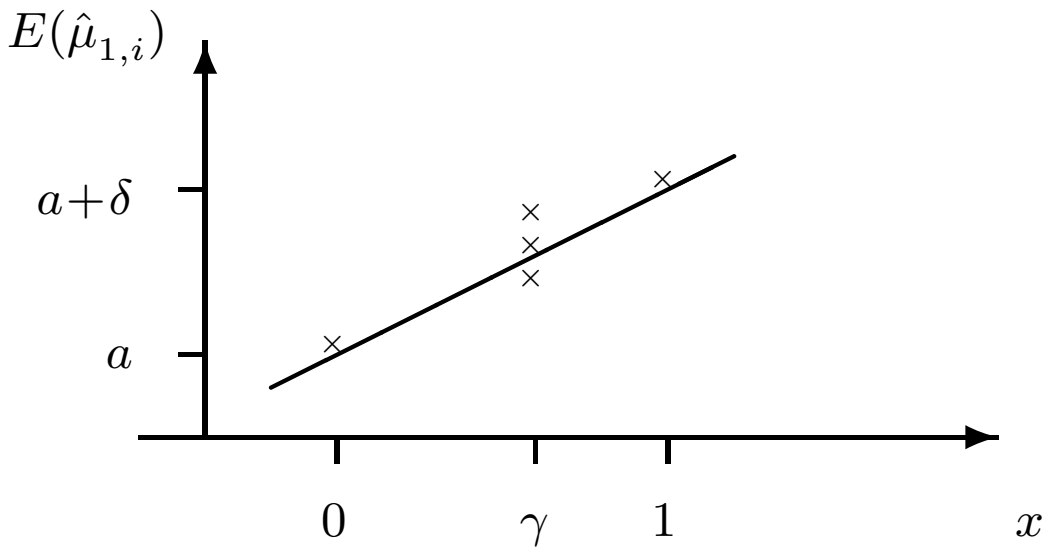
$\hat{\mu}_{1,c}$ has $x = 0$,

$\hat{\mu}_{1,1}, \ldots, \hat{\mu}_{1,k-1}$ have $x = \gamma$,

$\hat{\mu}_{1,k}$ has $x = 1$,

and we are interested in the slope δ.

Alternative derivation of MLEs in Phase II



It is easy to see that

for $\gamma = 0.5$, $\quad \hat{\delta} = \hat{\mu}_{1,k} - \hat{\mu}_{1,c} = \hat{\theta}_{1,k}$,

for $\gamma < 0.5$, $\quad \hat{\mu}_{1,1}, \ldots, \hat{\mu}_{1,k}$ contribute to $\hat{\delta}$ with negative weights,

for $\gamma > 0.5$, $\quad \hat{\mu}_{1,1}, \ldots, \hat{\mu}_{1,k}$ contribute to $\hat{\delta}$ with positive weights.