Warwick Medical School

Combining short-term and long-term endpoint data in a clinical trial with treatment selection

Cornelia Kunz[1], Nick Parsons[1], Sue Todd[2], Tim Friede[3] and Nigel Stallard[1*]

[1]University of Warwick, [2]University of Reading, [3]Universitaetsmedizin Göttingen

[*] *n.stallard@warwick.ac.uk*

THE UNIVERSITY OF WARWICK

Motivating example

A trial in Alzheimer's disease
- experimental treatment at 3 doses and placebo control

Primary endpoint
- ADAS-cog change over 12 weeks

We want to
- select most effective dose
- provide a valid comparison with control

Motivating example

A trial in Alzheimer's disease
- experimental treatment at 3 doses and placebo control

Primary endpoint
- ADAS-cog change over 12 weeks

We want to
- select most effective dose
- provide a valid comparison with control

We will focus on treatment selection

Treatment selection

Treatment selection

- often conducted in phase II(b) trial
- often based on short-term endpoint
  here ADAS-cog change over 6 weeks
- aims to select dose/treatment with best long-term endpoint

Treatment selection

Treatment selection
- often conducted in phase II(b) trial
- often based on short-term endpoint
  here ADAS-cog change over 6 weeks
- aims to select dose/treatment with best long-term endpoint

But may well have long-term endpoint data for some patients
Can we use this to improve selection?

Have short-term data for all and long-term data for some patients

Have short-term data for all and long-term data for some patients

Question 1

Which is best approach:

- use available long-term data only?
- use available short-term data only?
- use combination of long- and short-term data? How?

Have short-term data for all and long-term data for some patients

Question 1

Which is best approach:

- use available long-term data only?
- use available short-term data only?
- use combination of long- and short-term data? How?

Question 2

If we use short-term endpoint, what is best endpoint to use?

- E.g. could use eg 4 week data:
  more data, less correlated with long-term endpoint
- E.g. could use 8 week data:
  fewer data, more correlated with long-term endpoint

Short-term and long-term endpoint model

k doses (possibly + control)

Short-term endpoint (N/group), $X_{i,j}, i = 1, \ldots, k, j = 1, \ldots, N$
$X_{i,j} \sim N(\mu_{0i}, \sigma_0^2)$

Short-term and long-term endpoint model

k doses (possibly + control)

Short-term endpoint (N/group), $X_{i,j}, i = 1, \ldots, k, j = 1, \ldots, N$
$X_{i,j} \sim N(\mu_{0i}, \sigma_0^2)$

Long-term endpoint (n/group), $Y_{i,j}, i = 1, \ldots, k, j = 1, \ldots, n$
$Y_{i,j} \sim N(\mu_i, \sigma^2)$

Short-term and long-term endpoint model

k doses (possibly + control)

Short-term endpoint (N/group), $X_{i,j}, i = 1, \ldots, k, j = 1, \ldots, N$
$X_{i,j} \sim N(\mu_{0i}, \sigma_0^2)$

Long-term endpoint (n/group), $Y_{i,j}, i = 1, \ldots, k, j = 1, \ldots, n$
$Y_{i,j} \sim N(\mu_i, \sigma^2)$

$\text{corr}(X_{i,j}, Y_{i,j}) = \rho_w$
$\text{corr}(X_{i,j}, X_{i',j'}) = \text{corr}(Y_{i,j}, Y_{i',j'}) = 0, i \neq i'$ or $j \neq j'$

Short-term and long-term endpoint model

k doses (possibly + control)

Short-term endpoint (N/group), $X_{i,j}, i = 1, \ldots, k, j = 1, \ldots, N$
$X_{i,j} \sim N(\mu_{0i}, \sigma_0^2)$

Long-term endpoint (n/group), $Y_{i,j}, i = 1, \ldots, k, j = 1, \ldots, n$
$Y_{i,j} \sim N(\mu_i, \sigma^2)$

$\text{corr}(X_{i,j}, Y_{i,j}) = \rho_w$
$\text{corr}(X_{i,j}, X_{i',j'}) = \text{corr}(Y_{i,j}, Y_{i',j'}) = 0, i \neq i'$ or $j \neq j'$

We wish to select treatment with largest μ_i

Treatment selection

Using long-term endpoint data only

- Obtain estimate $\tilde{\mu}_i = \sum_{j=1}^{n} Y_{ij}/n, i = 1, \ldots, k$
- Select treatment with largest $\tilde{\mu}_i$

Treatment selection

Using long-term endpoint data only

- Obtain estimate $\tilde{\mu}_i = \sum_{j=1}^{n} Y_{ij}/n, i = 1, \ldots, k$
- Select treatment with largest $\tilde{\mu}_i$

Using short-term endpoint data only

- Obtain estimate $\hat{\mu}_{0i} = \sum_{j=1}^{N} X_{ij}/N, i = 1, \ldots, k$
- Select treatment with largest $\hat{\mu}_{0i}$

Treatment selection

Using long-term endpoint data only

- Obtain estimate $\tilde{\mu}_i = \sum_{j=1}^{n} Y_{ij}/n, i = 1, \ldots, k$
- Select treatment with largest $\tilde{\mu}_i$

Using short-term endpoint data only

- Obtain estimate $\hat{\mu}_{0i} = \sum_{j=1}^{N} X_{ij}/N, i = 1, \ldots, k$
- Select treatment with largest $\hat{\mu}_{0i}$

Using combination of short-term and long-term endpoint data

- Obtain estimate $\hat{\mu}_i$ from bivariate model

$$\begin{pmatrix} X_{i,j} \\ Y_{i,j} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_{0i} \\ \mu_i \end{pmatrix}, \begin{pmatrix} \sigma_0^2 & \rho_w \sigma \sigma_0 \\ \rho_w \sigma \sigma_0 & \sigma^2 \end{pmatrix}\right)$$

- Select treatment with largest $\hat{\mu}_i$

Treatment selection

$$\tilde{\mu}_i \sim N(\mu_i, \sigma^2/n)$$

$$\hat{\mu}_{0i} \sim N(\mu_{0i}, \sigma_0^2/N)$$

$$\hat{\mu}_i \sim N(\mu_i, \sigma^2/n^*)$$

where $n^* = \left(\frac{1}{n} - \rho_w^2 \left(\frac{1}{n} - \frac{1}{N}\right)\right)^{-1}$

Treatment selection

$$\tilde{\mu}_i \sim N(\mu_i, \sigma^2/n)$$

$$\hat{\mu}_{0i} \sim N(\mu_{0i}, \sigma_0^2/N)$$

$$\hat{\mu}_i \sim N(\mu_i, \sigma^2/n^*)$$

where $n^* = \left(\frac{1}{n} - \rho_w^2\left(\frac{1}{n} - \frac{1}{N}\right)\right)^{-1}$

Hence get probability of selecting certain treatment with each method for given μ_i, μ_{0i}, σ, σ_0, ρ_w

Gain from using short-term endpoint

n^* is 'equivalent sample size' from use of short-term information

$$n^* = \left(\frac{1}{n} - \rho_w^2 \left(\frac{1}{n} - \frac{1}{N} \right) \right)^{-1}$$

Note: $n^* \geq n$

Gain from using short-term endpoint

n^* is 'equivalent sample size' from use of short-term information

$$n^* = \left(\frac{1}{n} - \rho_w^2 \left(\frac{1}{n} - \frac{1}{N} \right) \right)^{-1}$$

Note: $n^* \geq n$

$\rho_w = 0 \implies n^* = n$

$\rho_w = \pm 1 \implies n^* = N$

Gain from using short-term endpoint

n^* is 'equivalent sample size' from use of short-term information

$$n^* = \left(\frac{1}{n} - \rho_w^2 \left(\frac{1}{n} - \frac{1}{N} \right) \right)^{-1}$$

Note: $n^* \geq n$

$\rho_w = 0 \implies n^* = n$

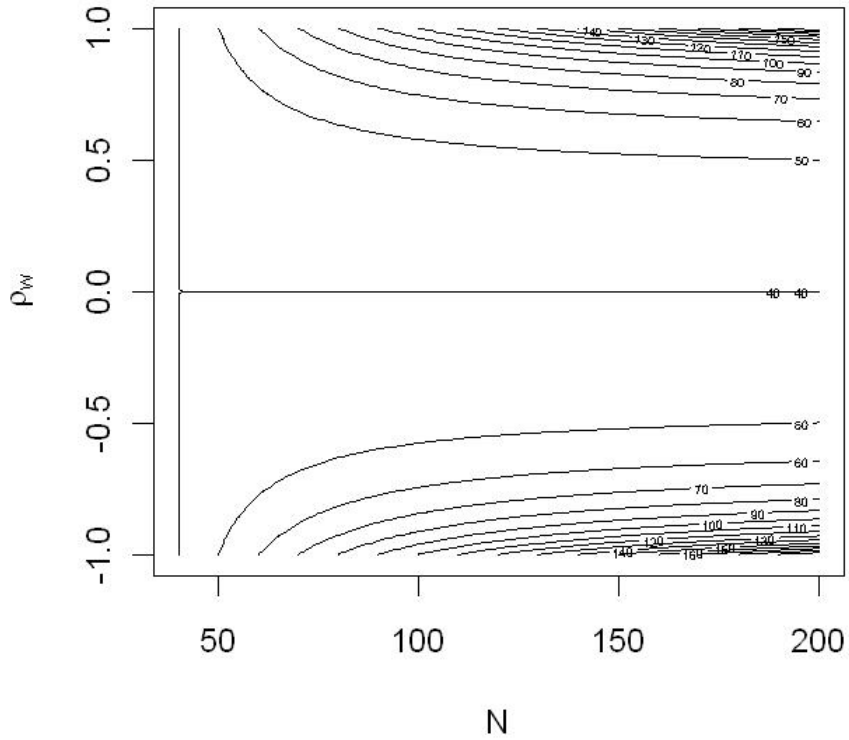
$\rho_w = \pm 1 \implies n^* = N$

Endpoint 1: ρ_w, N

Endpoint 2: ρ'_w, N'

Endpoint 1 is preferable if $\rho_w \left(\frac{1}{n} - \frac{1}{N} \right) > \rho'_w \left(\frac{1}{n} - \frac{1}{N'} \right)$

Trade-off between larger N and larger ρ_w (here with $n = 40$)



Random effects model

Assume

$$\begin{pmatrix} X_{i,j} \\ Y_{i,j} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_{0i} \\ \mu_i \end{pmatrix}, \begin{pmatrix} \sigma_0^2 & \rho_w \sigma \sigma_0 \\ \rho_w \sigma \sigma_0 & \sigma^2 \end{pmatrix}\right)$$

Random effects model

Assume

$$\begin{pmatrix} X_{i,j} \\ Y_{i,j} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_{0i} \\ \mu_i \end{pmatrix}, \begin{pmatrix} \sigma_0^2 & \rho_w\sigma\sigma_0 \\ \rho_w\sigma\sigma_0 & \sigma^2 \end{pmatrix}\right)$$

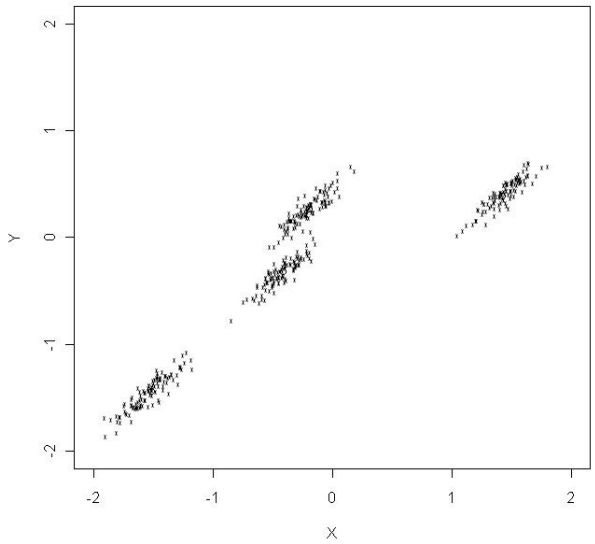
$$\begin{pmatrix} \mu_{0i} \\ \mu_i \end{pmatrix} \sim N\left(\begin{pmatrix} \theta_{0i} \\ \theta_i \end{pmatrix}, \begin{pmatrix} \tau_0^2 & \rho_b\tau\tau_0 \\ \rho_b\tau\tau_0 & \tau^2 \end{pmatrix}\right)$$

Note:

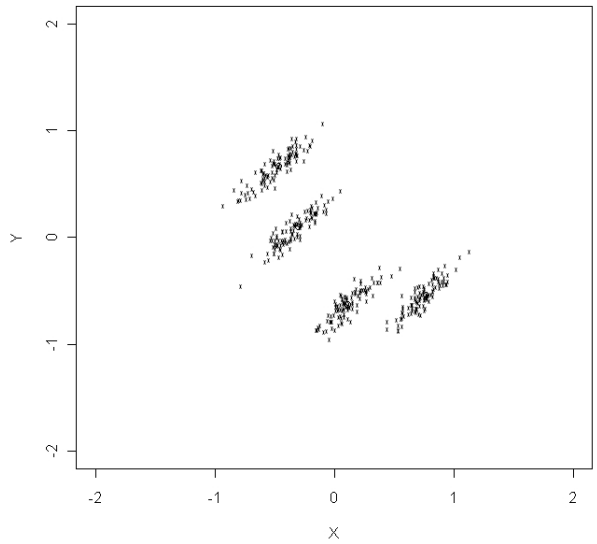
- we still wish to select treatment with largest μ_i (not θ_i)
- we will use fixed effects model to draw inference on specific treatments
- we will use random effects model to understand properties of different approaches

ρ_w and ρ_b

$\rho_w > 0, \rho_b > 0$



$\rho_w > 0, \rho_b < 0$



(here with $\sigma = \sigma_0 < \tau = \tau_0$ for clarity)

Note: for surrogate endpoint require both ρ_w and ρ_b are large

Treatment selection

Random effects model

$$\begin{pmatrix} \tilde{\mu}_i \\ \mu_i \end{pmatrix} \sim N\left(\begin{pmatrix} \theta_i \\ \theta_i \end{pmatrix}, \begin{pmatrix} \sigma^2/n + \tau^2 & \tau^2 \\ \tau^2 & \tau^2 \end{pmatrix}\right) \quad (1)$$

$$\begin{pmatrix} \hat{\mu}_{0i} \\ \mu_i \end{pmatrix} \sim N\left(\begin{pmatrix} \theta_{0i} \\ \theta_i \end{pmatrix}, \begin{pmatrix} \sigma_0^2/N + \tau_0^2 & \rho_b \tau_0 \tau \\ \rho_b \tau_0 \tau & \tau^2 \end{pmatrix}\right) \quad (2)$$

$$\begin{pmatrix} \hat{\mu}_i \\ \mu_i \end{pmatrix} \sim N\left(\begin{pmatrix} \theta_i \\ \theta_i \end{pmatrix}, \begin{pmatrix} \sigma^2/n^* + \tau^2 & \tau^2 \\ \tau^2 & \tau^2 \end{pmatrix}\right) \quad (3)$$

hence get probability that

$\arg\max_{i=1,\ldots,k}\{\hat{\mu}_{0i}\} = \arg\max_{i=1,\ldots,k}\{\mu_i\}$ etc.

Treatment selection

Random effects model

$$\begin{pmatrix} \tilde{\mu}_i \\ \mu_i \end{pmatrix} \sim N\left(\begin{pmatrix} \theta_i \\ \theta_i \end{pmatrix}, \begin{pmatrix} \sigma^2/n + \tau^2 & \tau^2 \\ \tau^2 & \tau^2 \end{pmatrix}\right) \quad (1)$$

$$\begin{pmatrix} \hat{\mu}_{0i} \\ \mu_i \end{pmatrix} \sim N\left(\begin{pmatrix} \theta_{0i} \\ \theta_i \end{pmatrix}, \begin{pmatrix} \sigma_0^2/N + \tau_0^2 & \rho_b \tau_0 \tau \\ \rho_b \tau_0 \tau & \tau^2 \end{pmatrix}\right) \quad (2)$$

$$\begin{pmatrix} \hat{\mu}_i \\ \mu_i \end{pmatrix} \sim N\left(\begin{pmatrix} \theta_i \\ \theta_i \end{pmatrix}, \begin{pmatrix} \sigma^2/n^* + \tau^2 & \tau^2 \\ \tau^2 & \tau^2 \end{pmatrix}\right) \quad (3)$$

hence get probability that

$\arg\max_{i=1,\ldots,k}\{\hat{\mu}_{0i}\} = \arg\max_{i=1,\ldots,k}\{\mu_i\}$ etc.

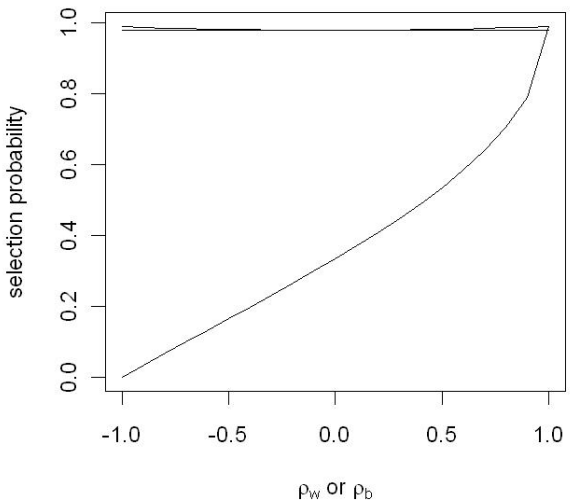
Note:

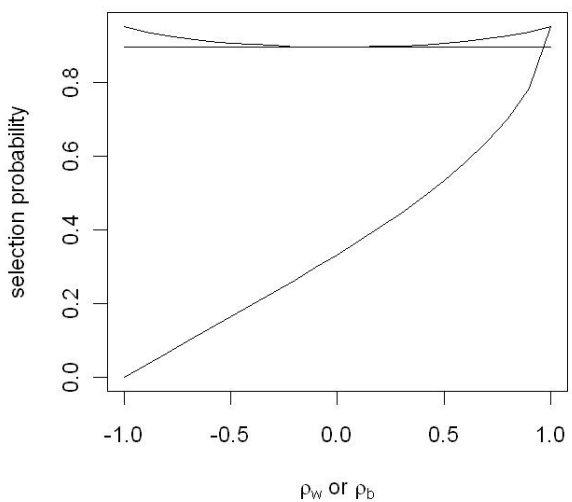
(2) depends on ρ_b, but not on ρ_w

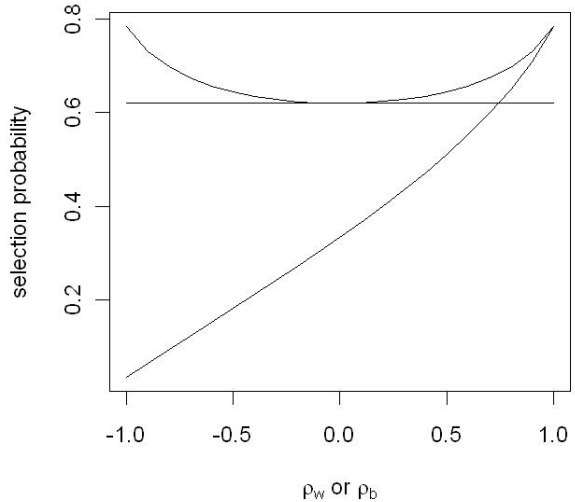
(3) depends on ρ_w (actually on ρ_w^2) via n^*, but not on ρ_b

Example: $k = 3, n = 20, N = 100, \theta_1 = \theta_2 = \theta_3, \theta_{01} = \theta_{02} = \theta_{03}$

$5\sigma = 5\sigma_0 = \tau = \tau_0$ $\sigma = \sigma_0 = \tau = \tau_0$ $\sigma = \sigma_0 = 5\tau = 5\tau_0$

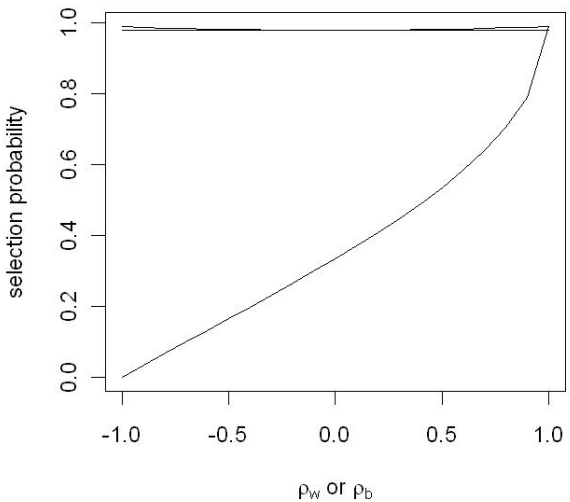


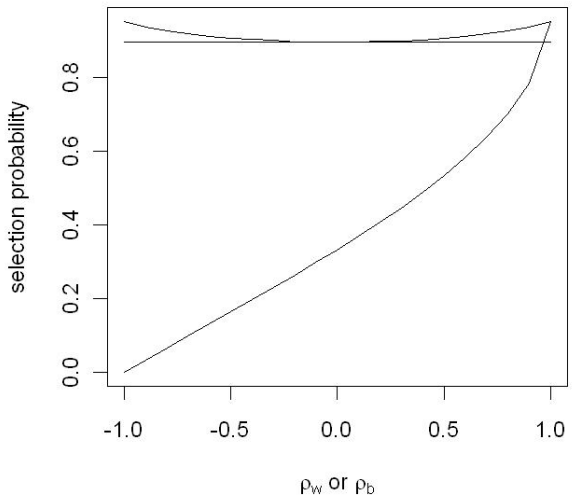


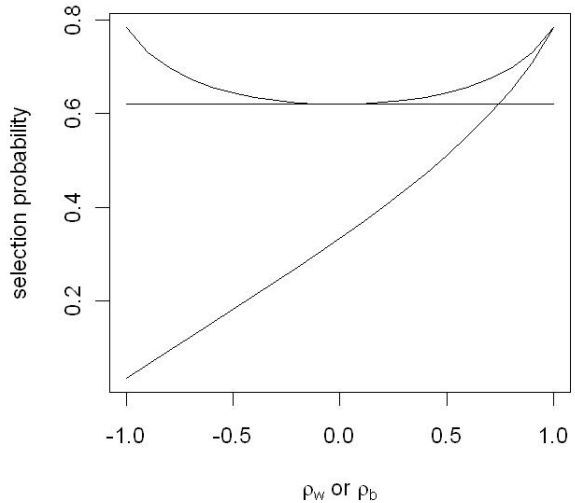


Example: $k = 3, n = 20, N = 100, \theta_1 = \theta_2 = \theta_3, \theta_{01} = \theta_{02} = \theta_{03}$

$5\sigma = 5\sigma_0 = \tau = \tau_0$ $\sigma = \sigma_0 = \tau = \tau_0$ $\sigma = \sigma_0 = 5\tau = 5\tau_0$



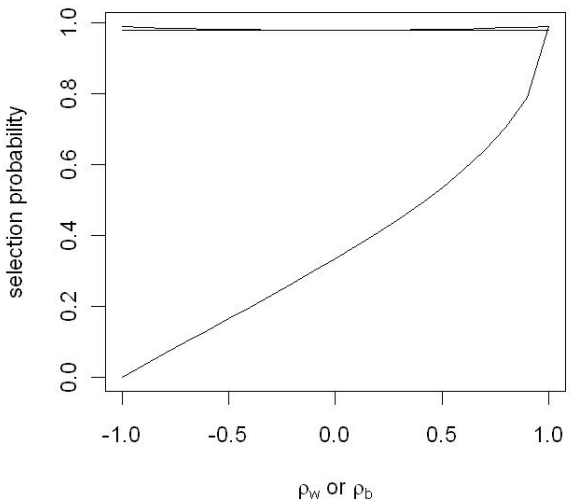


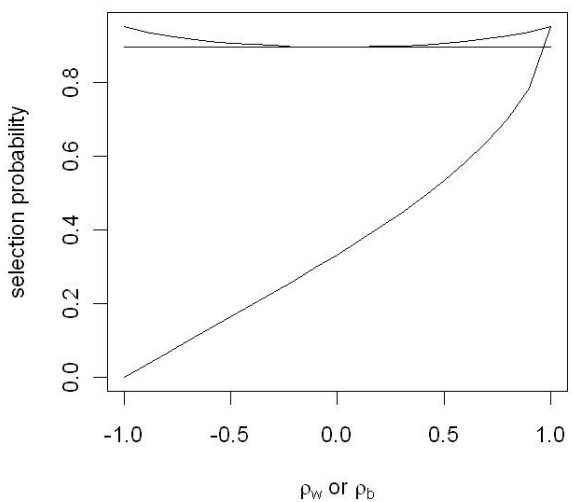


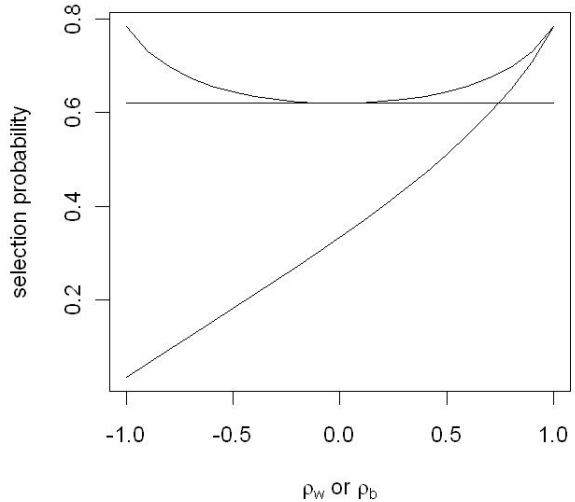
Using short-term endpoint data alone can be better
- if ρ_b large and $\rho_b > \rho_w$, particularly if τ and τ_0 are small

Example: $k = 3, n = 20, N = 100, \theta_1 = \theta_2 = \theta_3, \theta_{01} = \theta_{02} = \theta_{03}$

$5\sigma = 5\sigma_0 = \tau = \tau_0$ $\sigma = \sigma_0 = \tau = \tau_0$ $\sigma = \sigma_0 = 5\tau = 5\tau_0$







Using short-term endpoint data alone can be better
- if ρ_b large and $\rho_b > \rho_w$, particularly if τ and τ_0 are small

Using short-term endpoint data alone can be (a lot) worse
- if ρ_b is small or negative

Conclusions

Using short-term endpoint data (on more patients) with long-term endpoint data always improves precision and treatment selection

- Gain small for small ρ_w, but can be large for large ρ_w or N
- Choice of best short-term endpoint depends on $\rho_w(1/n - 1/N)$

Conclusions

Using short-term endpoint data (on more patients) with long-term endpoint data always improves precision and treatment selection
- Gain small for small ρ_w, but can be large for large ρ_w or N
- Choice of best short-term endpoint depends on $\rho_w(1/n - 1/N)$

Using short-term endpoint data alone can improve treatment selection
- if ρ_b is large and $\rho_b > \rho_w$, particularly if τ and τ_0 are small

Using short-term endpoint data alone can make treatment selection worse
- if ρ_b is small or negative

Extensions

Adaptive design that allows combination of treatment selection phase with confirmatory two-arm phase
- final analysis combines data from both phases and controls
(familywise) type I error rate

Data-driven selection method; uses estimates of $\rho_w, \rho_b, \sigma, \sigma_0, \tau, \tau_0$ to choose whether to select based on $\hat{\mu}_{0i}$ or $\hat{\mu}_i$
- in most cases does as well as better of two methods
- in some cases does better than either method
- may be able to find better method