

Multiplicity and Estimation

P.Bauer

Medical University of Vienna

London, November 2012

- Selection bias
- Reporting bias
- Bias at admission
- Methods of estimation to reduce bias
- Multiple testing – Multiple confidence intervals

The first scenario to be considered

- **To compare k treatments with a single control**
Independent normal distributions, equal known variance σ^2 , means $\mu_1, \dots, \mu_k$ and μ_0 , respectively
- The same sample size n is planned in all groups
- Planned interim analysis after a fraction of rn , $0 \leq r \leq 1$, observations in every group
- **The best treatment (and the control) are selected and investigated at the second stage**
- Quantify mean bias and mean square error (MSE) of the conventional ML estimates of the mean treatment to control differences

The conventional fixed sample size design ($r=1$)

- It is correct that the k final treatment vs control effect estimates are unbiased
- However, it would be hiding ones head in the sand to ignore that the magnitude of the effects plays an important role in decisions and actions following such a trial
- **E.g., the plausible strategy to go on with the most effective (and sufficiently safe) dose will tend to produce positively biased estimates of the true effect size of this dose in planning the next steps of drug development**

Notation

$\bar{X}_i, i = 0, 1, \dots, k$, means from the rn observations at the first stage

$\bar{Y}_j$ corresponding mean from the $(1-r)n$ observations at the second stage, given that treatment j has been selected, the control ($i = 0$) is always selected

$\bar{Z}_s = r\bar{X}_s + (1-r)\bar{Y}_s$ the mean of a selected treatment in the final analysis

$\bar{Z}_s - \bar{Z}_0$ final treatment effect estimate for the selected treatment

Selecting the best treatment

Selection bias

DAHIYA, JASA, 1974; POSCH et al., *Stat Med*, 2005

$$b_{sel}(\hat{\delta}_s) = b_{sel}(\bar{Z}_s - \bar{Z}_0) = \\ \sum_{j=1}^k E[\bar{Z}_j - \mu_j \mid \bar{X}_j = \bar{X}_{[k]}] P(\bar{X}_j = \bar{X}_{[k]}),$$

This holds because $\bar{Z}_0$ is an unbiased estimate of μ_0

Selecting the best treatment

Mean square error

The selection mean square error

$$MSE_{sel}(\hat{\delta}_s) = \sum_{j=1}^k E[(\bar{Z}_j - \bar{Z}_0 - (\mu_j - \mu_0))^2 \mid \dots]$$

can be defined accordingly, however, the variability arising from the mean of the control group has to be accounted for

Selecting the best treatment

Maximum bias

The selection bias is largest if all the treatment means are equal ($\mu_1 = \mu_2 = \dots = \mu_k$)

Proof for $k=2$:

PUTTER & RUBINSTEIN, *Technical Report TR 165, Statistics Department, University of Wisconsin, 1968.*

STALLARD, TODD & WHITEHEAD, *JSPI, 2008.*

For $k=3$: Numerical solution in

BAUER et al., *Stat Med, 2009*

General proof:

CARRERAS & BRANNATH, *Stat Med, 2012*

Selecting the best treatment

Maximum bias and MSE

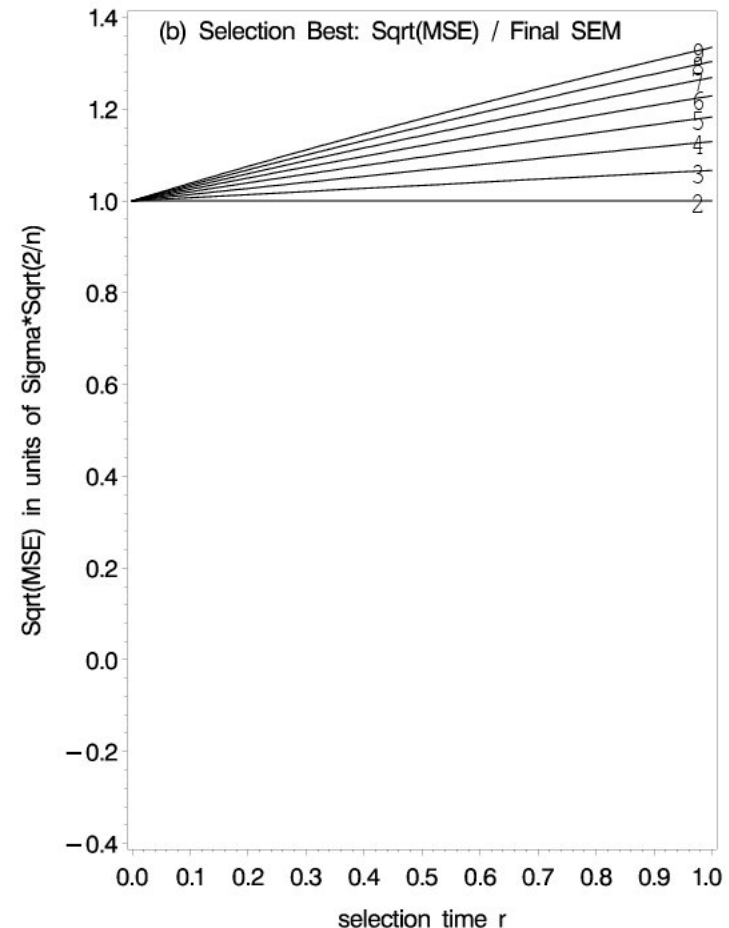
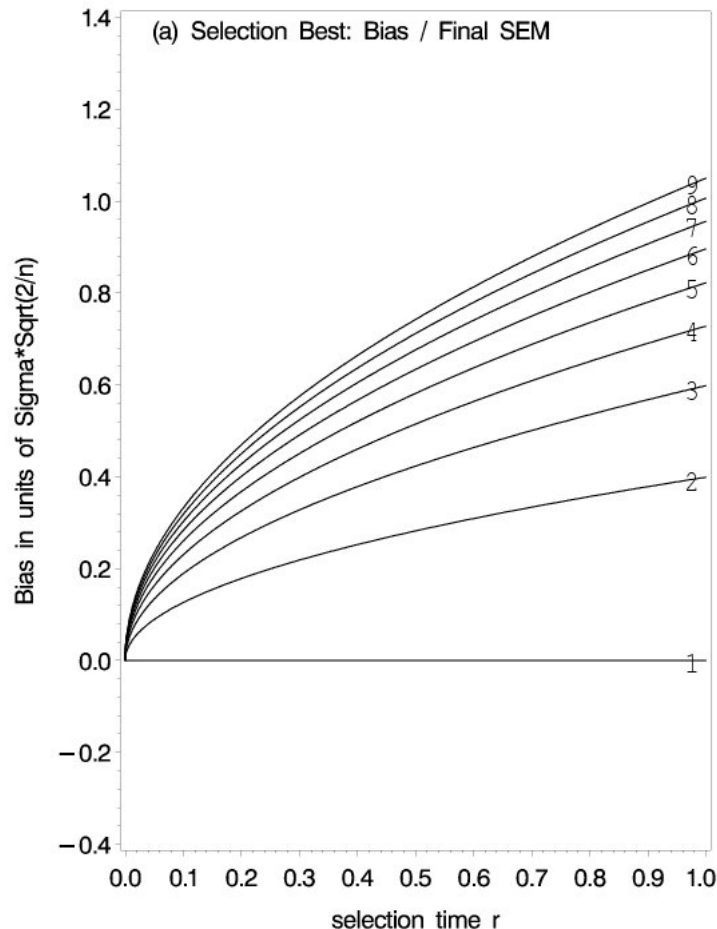
Under the „worst case scenario“ of equal treatment means closed formula for bias and MSE can be derived (P.BAUER, et al., *Stat Med*, 2009):

$$b_{sel}(\hat{\delta}_{[k]}) = m_1(k) \sqrt{r \sigma^2 / n}$$

$$MSE_{sel}(\hat{\delta}_{[k]}) = (\sigma^2 / n) \{[m_2(k) - 1]r + 2\}$$

where $m_1(k)$ and $m_2(k)$ are the first and second moment of the maximum of k independent standard normal random variables.

Maximum mean selection bias and $\sqrt{MSE}$ (both in units of $\sigma\sqrt{(2/n)}$) as a function of k and r Selecting the best



To take home

- Random selection of a treatment ($r=0$) - no bias
- The (maximum) bias increases with increasing number of treatments k , tends to infinity for $k \rightarrow \infty$
- It sharply increases with r and is largest for $r=1$ („post trial selection“) !
- However, for differing treatment means earlier selection increases the probability of wrong selections due to the larger variability
- If a treatment is considerably better than the others the bias decreases with the margin since the probability of being selected increases, in the limit the estimate is unbiased with conventional MSE

To take home (cont.)

- The corresponding $\sqrt{MSE}$ does not increase with k to the same extent as the bias
- It is identical for $k=2$ and $k=1$ which holds true under some general symmetry conditions
POSCH et al., *Stat Med*, 2005
- In units of the conventional standard error at the end $\sqrt{MSE}$ increases close to linear with the “selection time” r

Reporting bias (selecting the best treatment)

- Each observed effect estimate is reported separately regardless of selection
- We report the effect estimate in the total sample if the treatment has been selected and the interim effect estimate if it has not been selected

$$b_{rep}(\hat{\delta}_j) = E[\bar{Z}_j - \mu_j \mid \bar{X}_j = \bar{X}_{[k]}]P(\bar{X}_j = \bar{X}_{[k]}) + \\ E[\bar{X}_j - \mu_j \mid \bar{X}_j < \bar{X}_{[k]}]P(\bar{X}_j < \bar{X}_{[k]})$$

The reporting *MSE* can be defined accordingly!

Reporting bias - to take home

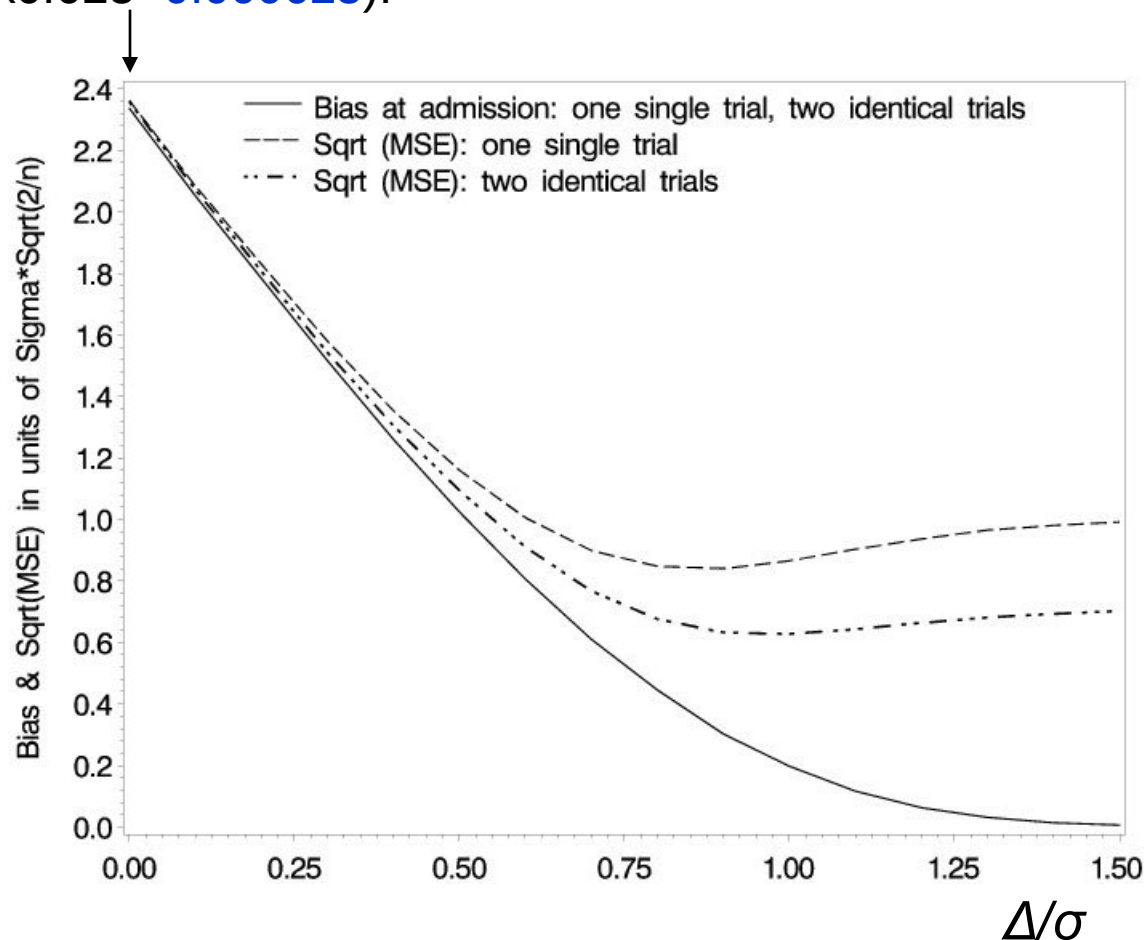
- For **equal treatment means** the reporting bias **generally is negative**: On the one hand if the interim effect is large we tend to dilute the treatment effect by the independent second sample. On the other hand if the interim effect is small we tend to stay with the small effect as it is
- It is most accentuated and **equal** for $k=2$ and $k=3$
- As k increases the probability to be selected decreases. For any j we more often will use the hardly biased first stage estimate, the reporting bias coming closer to zero
- For $r \rightarrow 1$ (no selection) the reporting bias tends to zero
- For $r \rightarrow 0$ a treatment is selected with a highly variable effect estimate whose distribution is shifted to the left (the reporting bias diverges to minus infinity)

Interlude: admission bias

- **Example:** Two identical independent trials comparing a new treatment to a control
- Each of the preplanned one sided z-tests for the primary outcome variable at the level 0.025 has a power of 90% at an effect size of $\Delta/\sigma=1$
- Estimates are only reported (or relevant for the public in case of registration of a new drug) if both one sided z-tests have been rejected!
- **This will result in a “bias at admission”**
- See earlier work on bias in meta-analyses:
HEDGES, *J.Educat.Stat.*, 1984; HEDGES & OLKIN, 1985; BEGG & BERLIN, *J.R.S.S.A*, 1988

Admission bias (one or two pivotal trials) as a function of the true effect size Δ/σ

Here the probability for registration
is small ($0.025 \times 0.025 = 0.000625$)!



Admission bias – to take home

- The mean bias is largest for $\Delta=0$. However, here rejection only occurs with a probability of 0.025 or 0.000625 (for two independent trials)
- It may be quite substantial for lower effect sizes
- **If the true effect is close to the targeted effect size the bias is small and for increasing effect sizes approaches 0 quite fast**
- The bias is equal for the single or two studies scenario
- The *MSE* is lower in the two studies scenario.
- If in the single study scenario the true effect size is slightly below the targeted one the *MSE* is slightly below the conventional mean square (**truncated distribution!**)

Methods of estimation to reduce bias

$$\tilde{\delta} = \overset{\text{MLE}}{\hat{\delta}} - \overset{\text{Bias}}{b(\tilde{\delta})}$$

Equation for the **bias corrected estimator** solved by numerical iteration [created for sequential trials]

$$\text{Bias}(\tilde{\delta}) = E(\tilde{\delta}) - \delta = [E(\hat{\delta}) - \delta] - E[b(\tilde{\delta})] = b(\delta) - \overset{\text{bias correction}}{E[b(\tilde{\delta})]}$$

WHITEHEAD, *Biometrika* (1986)

Extended to correction of the ML-estimate when the best treatment has been selected at interim by

STALLARD and TODD, *JSPI* (2005)

Shrinkage estimators

In conventional multiarmed trials:

$$\dot{\delta}_s = \hat{C}_+ \hat{\delta}_s + (1 - \hat{C}_+) \bar{\delta}, \text{ where } \hat{C}_+ = \max(\hat{C}, 0), \hat{C} = 1 - \frac{f(k)\sigma^2}{n \sum_{j=1}^k (\hat{\delta}_j - \bar{\delta})^2}$$

← overall mean

LINDLEY *JRSS B* (1962), **HWANG** *Ind J Stat* (1993)

Extended to two stage designs with selection by

CARRERAS and **BRANNATH**, *Stat Med* (2012)

For the Unif. Min. Var. Cond. Unb. Estimate only the discrepancy between the largest and second largest mean at interim triggers the shrinkage of the MLE

COHEN and **SACKROWITZ**, *Statistics & Probability Letters* (1989)

BOWDEN and **GLIMM**, *Biom J* (2008)

Bias correction – to take home

- How to define bias in case of selection?
- What matters, bias or mean square error (or ...)?
- What is a suitable criterion for a „good“ estimate?
- Would we always also report conventional estimates - would regulators ask for it anyway?
- Correction and shrinkage becomes larger the more similar the effect estimates
- Should bias adjusted estimates be given in the spirit of a sensitivity analysis?

Simultaneous confidence intervals – stepwise multiple tests

KIM et al., *Statistical Decision Theory and Related Topics*, vol. IV (1988),, **STRASSBURGER** and **BRETZ**, *Stat Med* (2008)

- Contrary to conventional non-stepwise tests, for stepwise multiple tests (as the HOLM procedure) compatible simultaneous confidence intervals (**rejection $\equiv$ non-coverage**) are not straight forward
- The form depends on the choice of the set of hypotheses
- There might be situations where such a special choice is advantageous, e.g., two (opposite) one sided tests at level α each to show bioequivalence for the price that the corresponding confidence interval always covers the point of equivalence, **BAUER** and **KIESER**, *Biometrika* (1996)

Simultaneous confidence intervals (cont.)

- Compatible simultaneous (one-sided) stepwise CIs are ending at the null hypothesis (e.g., $(0, \infty)$ for the i^{th} “significant” treatment effect) thus only reflecting the stepwise decision
- The concept that such (new) simultaneous CIs only add information when the corresponding null hypothesis is **not** rejected in the multiple test is questionable, particularly when the rejected null hypotheses are the ones which are triggering further actions (as usual in the regulatory context)

Simultaneous confidence intervals (cont.)

- The focus on yes-or-no decisions may be the consequence of the far reaching yes-or-no decisions to be taken in regulatory processes
- The compatibility requirement should not be sacrificed to avoid further “multiplicity” (and ambiguity) – *personal opinion*
- **Does this question a lot of sophisticated work on simultaneous confidence intervals (or even on multiple tests) to be applied in other areas?**

Note, however, that the phenomenon of bias has to be considered as an intrinsic feature of human life when selecting, e.g., jobs, friends and partners based on a comparison of past observations afflicted by random variation

Thank you for your patience!