Statistical Inference following Self–Designing Clinical Trials with Binary Response

Wibke Stansen and Joachim Hartung

Department of Statistics, University of Dortmund

51th Annual Meeting of the German Medical Informatics, Biometry and Epidemiology Society,

10–14 September 2006, Leipzig

Outline

Self–Designing Clinical Trials

- introduction of "self–designing clinical trials" by Fisher (1998, Statist. Med.) for general setting of normal variables with known variances
- in Hartung (2001, Contr. Clin. Trials; 2006, Biom. J.) the distributional restriction is lifted by using a combining method of p–values
  - inverse normal method
- adaptive choice of both sample sizes and weights of the several study parts
- the rejection of the null hypothesis is tested just once at the end of the trial

Of Interest

- consider for a real valued parameter θ the test problem

$$H_0 : \theta = 0 \quad \text{versus} \quad H_1 : \theta > 0$$

  - test of level $\alpha/2$
  - confidence interval for θ of level $1 - \alpha$

- study is performed in K study parts where K is a finite random variable

Inverse Normal Method

in each step k, $k = 1, ..., K$:

- $\hat{\theta}_k$ unbiased estimator of θ
- test statistic T_k for testing H_0 vs. H_1
  assumption:
  - T_k is continuously distributed, otherwise approximative (binary case in detail later)
  - $T_k = T_k(\hat{\theta}_k)$ is (strictly) monotone increasing in $\hat{\theta}_k$

  $\hookrightarrow$ $T_k(\hat{\theta}_k - \theta)$ is (strictly) monotone decreasing in θ
- p-value $p_k = p_k(\theta) = 1 - F_{H_0}(T_k(\hat{\theta}_k - \theta))$
- transformation $z_k = \Phi^{-1}(1 - p_k) \sim N(0, 1)$ for true θ

Inverse Normal Method

- defining a sequence of nonnegative weights $w_1, ..., w_k, ...$ adaptively:
$$w_k = \hat{w}\{stage(0), ..., stage(k-1)\}$$

- with probability one under H_0 there exists a finite (random) K with
$$\sum_{k=1}^{K} w_k^2 = \sum_{k=1}^{\infty} w_k^2 = 1$$
then
$$Z_K = \sum_{k=1}^{K} w_k \, z_k = \sum_{k=1}^{\infty} w_k \, z_k \sim N(0,1) \quad \text{for true } \theta$$

- **decision rule:** H_0 is rejected at level $\frac{\alpha}{2}$ if $Z_{K|\theta=0} > \Phi^{-1}\left(1-\frac{\alpha}{2}\right)$

Practical Aspects

- specification of a lower bound for the weight of stage k
  $\Longrightarrow$ thus maximal number of stages is bounded

- also useful: specification of a minimal and maximal number of patients per stage

- during the course of the study design adaptions are possible and at every stage the next can be planned as the last one

- real planned studies for instance:
  - breast cancer study
  - Parkinson's disease study

Comment: fixing the weights a priori

$\Longrightarrow$ nearly an adaptive group sequential design of O'Brien and Fleming type (see Hartung, 2006, Biom. J.)

Overall p-Value

- overall p-value at trial termination:

$$p(\theta) = 1 - \Phi(Z_K(\theta))$$

  - $p(\theta)$ is a pivotal quantity increasing in θ
  - $p(\theta)$ follows an uniform distribution F on [0,1]

Confidence Interval and Point Estimator

construction of an $(1-\alpha)$–confidence interval for θ at the end of the trial

(see also: Liu and Chi, 2001; Wassmer, 2003; Hartung and Knapp, 2006)

- lower and upper bound:

$$\hat{\theta}_L = p^{-1}(\alpha/2) \quad \text{and} \quad \hat{\theta}_U = p^{-1}(1-\alpha/2)$$

- midpoint of the confidence interval:

$$\hat{\theta}_{1/2} = p^{-1}(1/2)$$

$\hookrightarrow$ median unbiased estimator for θ

Binary Outcomes

- parallel group design with

$$X_1 \sim B(n_1, p_1) \quad \text{and} \quad X_2 \sim B(n_2, p_2)$$

- parameters of interest:
  - risk difference: $D = p_1 - p_2$
  - logarithmic risk ratio: $\log RR = \log(p_1/p_2)$
  - logarithmic odds ratio: $\log OR = \log\left(\frac{p_1/(1-p_1)}{p_2/(1-p_2)}\right)$

Notation

fourfold table at stage k

treatment	success	failure	total
1	$n_{11,k}$	$n_{12,k}$	n_{1k}
2	$n_{21,k}$	$n_{22,k}$	n_{2k}
	$n_{11,k} + n_{21,k}$	$n_{12,k} + n_{22,k}$	n_k

Risk Difference

- estimation of $D = p_1 - p_2$ at stage k:

$$\hat{D}_k = \hat{p}_{1k} - \hat{p}_{2k} = \frac{n_{11,k}}{n_{1k}} - \frac{n_{21,k}}{n_{2k}}$$

- estimation of the variance of $\hat{D}_k$ at stage k:

$$\widehat{Var}(\hat{D}_k) = \widehat{Var}(\hat{p}_{1k}) + \widehat{Var}(\hat{p}_{2k}) = \frac{\hat{p}_{1k}(1-\hat{p}_{1k})}{n_{1k}-1} + \frac{\hat{p}_{2k}(1-\hat{p}_{2k})}{n_{2k}-1}$$

Logarithmic Risk Ratio

- estimation of $\log RR = \log(p_1/p_2)$ at stage k:

$$\log \widehat{RR}_k = \log\left(\hat{p}_{1k}/\hat{p}_{2k}\right) = \log\left(\frac{n_{11,k}/n_{1k}}{n_{21,k}/n_{2k}}\right)$$

- estimation of the variance of $\log \widehat{RR}_k$ at stage k:

$$\widehat{Var}(\log \widehat{RR}_k) = \frac{1}{n_{11,k}} - \frac{1}{n_{1k}} + \frac{1}{n_{21,k}} - \frac{1}{n_{2k}}$$

Logarithmic Odds Ratio

- estimation of $\log OR = \log((p_1/(1-p_1))/(p_2/(1-p_2)))$ at stage k:

$$\log \widehat{OR}_k = \log\left(\frac{\hat{p}_{1k}/(1-\hat{p}_{1k})}{\hat{p}_{2k}/(1-\hat{p}_{2k})}\right) = \log\left(\frac{n_{11,k}\ n_{22,k}}{n_{12,k}\ n_{21,k}}\right)$$

- estimation of the variance of $\log \widehat{OR}_k$ at stage k:

$$\widehat{Var}(\log \widehat{OR}_k) = \frac{1}{n_{11,k}} + \frac{1}{n_{12,k}} + \frac{1}{n_{21,k}} + \frac{1}{n_{22,k}}$$

Characteristics of the Test Statistic

- Test Statistic at stage k:

$$T_k = \frac{\hat{\theta}_k - \theta}{\sqrt{\widehat{Var}(\hat{\theta}_k)}} \overset{appr.}{\sim} N(0,1)$$

- $\frac{\partial\, T_k}{\partial\, \theta} = \frac{-1}{\sqrt{\cdots}} < 0 \Longrightarrow T_k$ is strictly monotone decreasing in θ

Example: Logarithmic Risk Ratio

$\alpha = 0.05$

stage k	sample size per group	weight w_k	log risk ratio	p–value p_k
1	20	0.447	0.368	0.108
2	63	0.559	0.288	0.024
3	56	0.698	0.381	0.009

test decision:

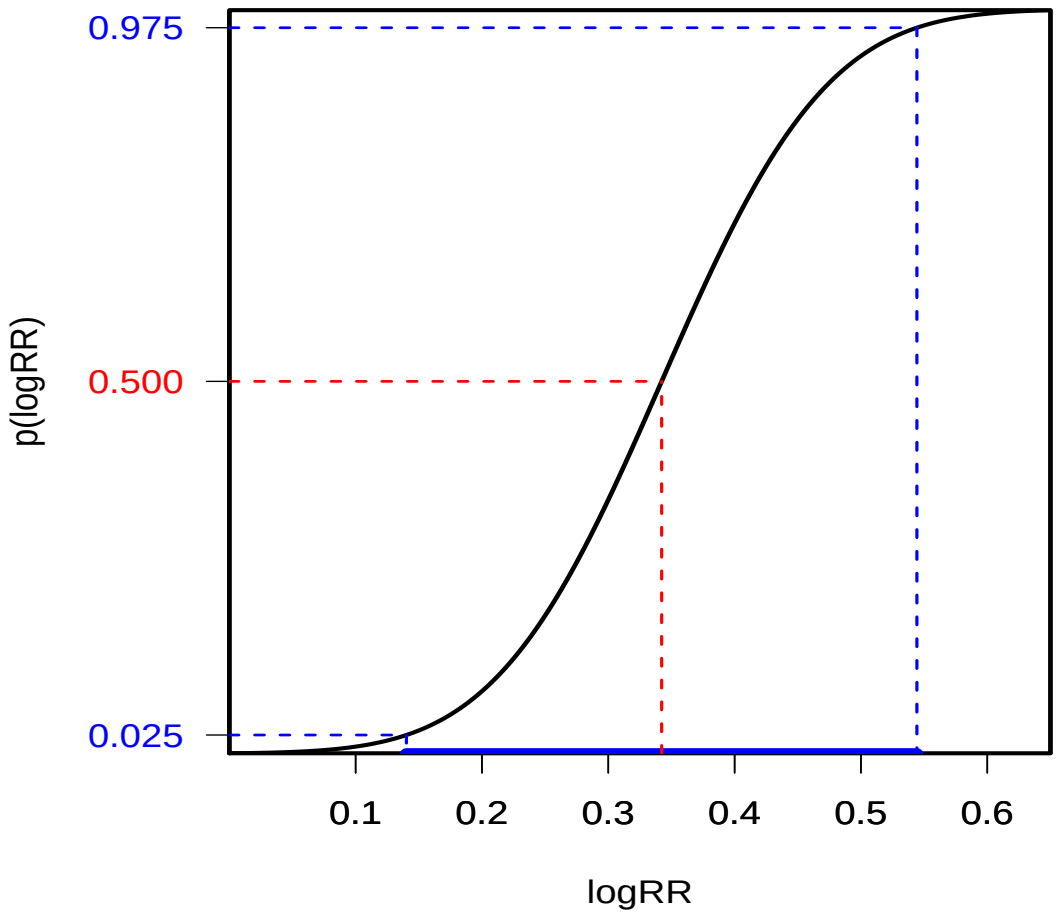
$$Z_{3|\log RR=0} = 3.320 > \Phi^{-1}(1 - 0.025) \Longrightarrow \text{rejection of } H_0$$

after trial termination:

- estimation of $\log RR = \log(p_1/p_2)$
- construction of an 95%-confidence interval for $\log RR$

Example: Logarithmic Risk Ratio

overall p–value: $p(\log RR) = 1 - \Phi(Z_K(\log RR))$



$$\log \widehat{RR}_{1/2} = 0.342$$

$$KI(\log RR) = [0.140, 0.544]$$

Some Simulation Results

- construction of a confidence interval with confidence level 95% and median unbiased estimation of D, $\log RR$ and $\log OR$

  $\hookrightarrow$ investigation of the

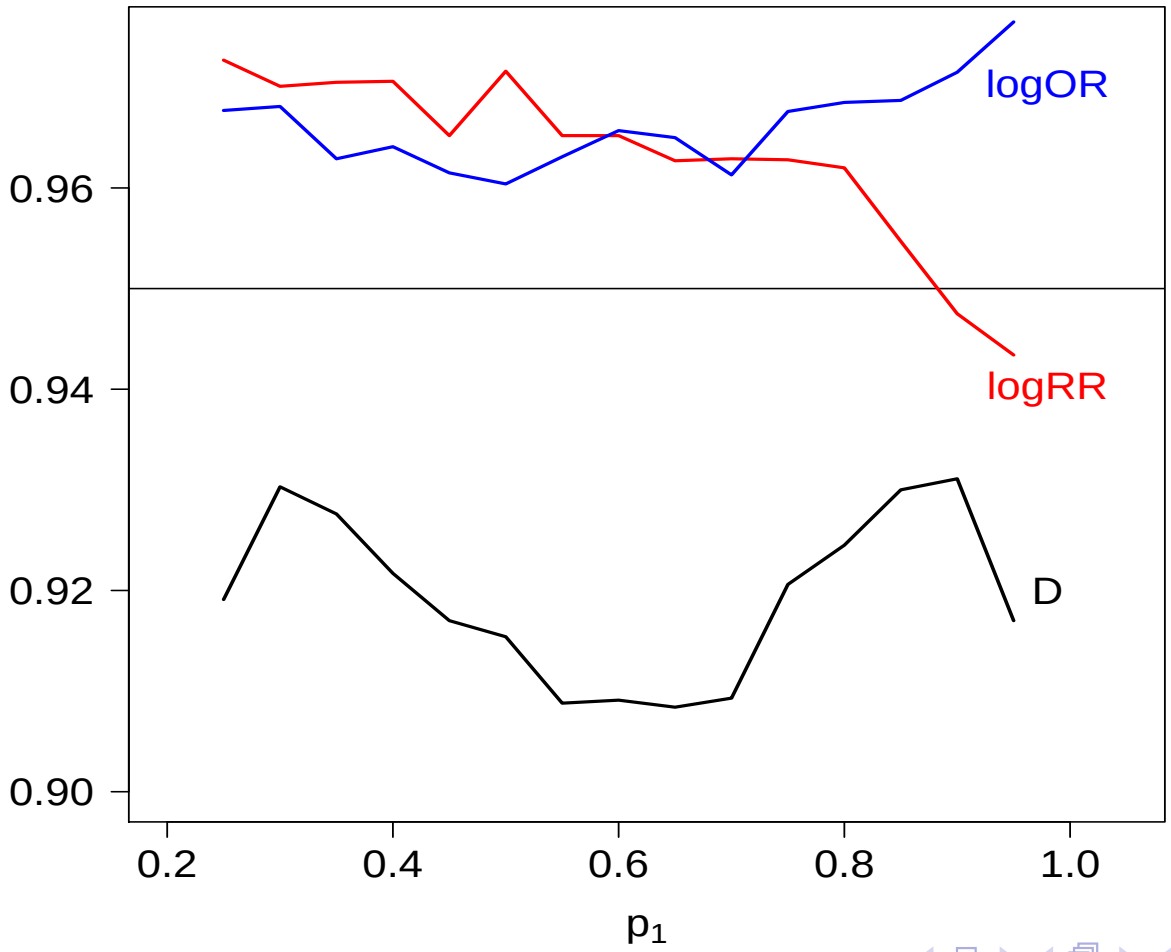
  - coverage probability of the confidence interval
  - average length of the confidence interval
  - point estimation

- adaptive choices of sample sizes and weights:
  learning rules of Hartung (2001, Contr. Clin. Trials)

- number of realized stages: *range*: $1 \leq K \leq 6$, *mean*: $\overline{K} \approx 3$

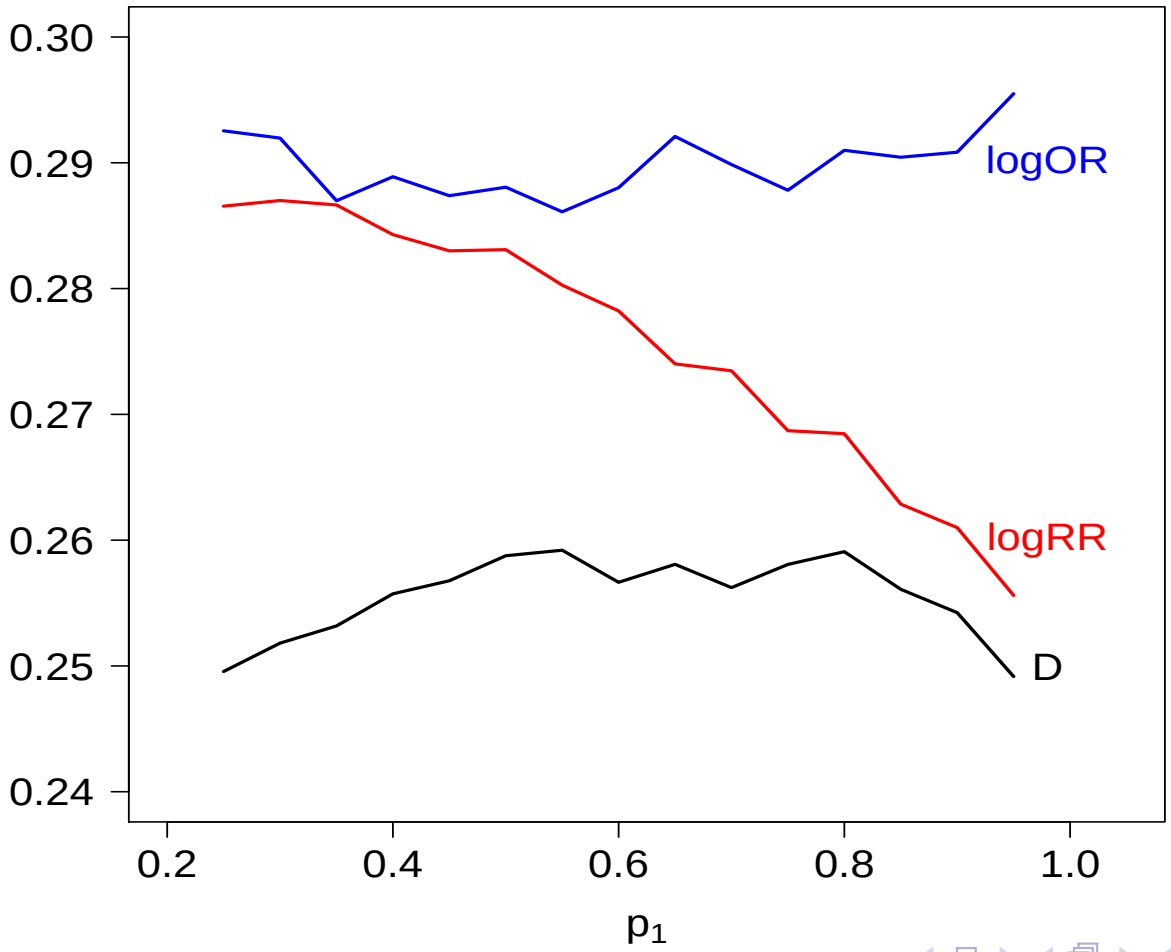
Coverage Probability of the Confidence Interval

- $D = \log RR = \log OR = 0.2$ fixed, $p_1 = 0.25, 0.3, ..., 0.95$



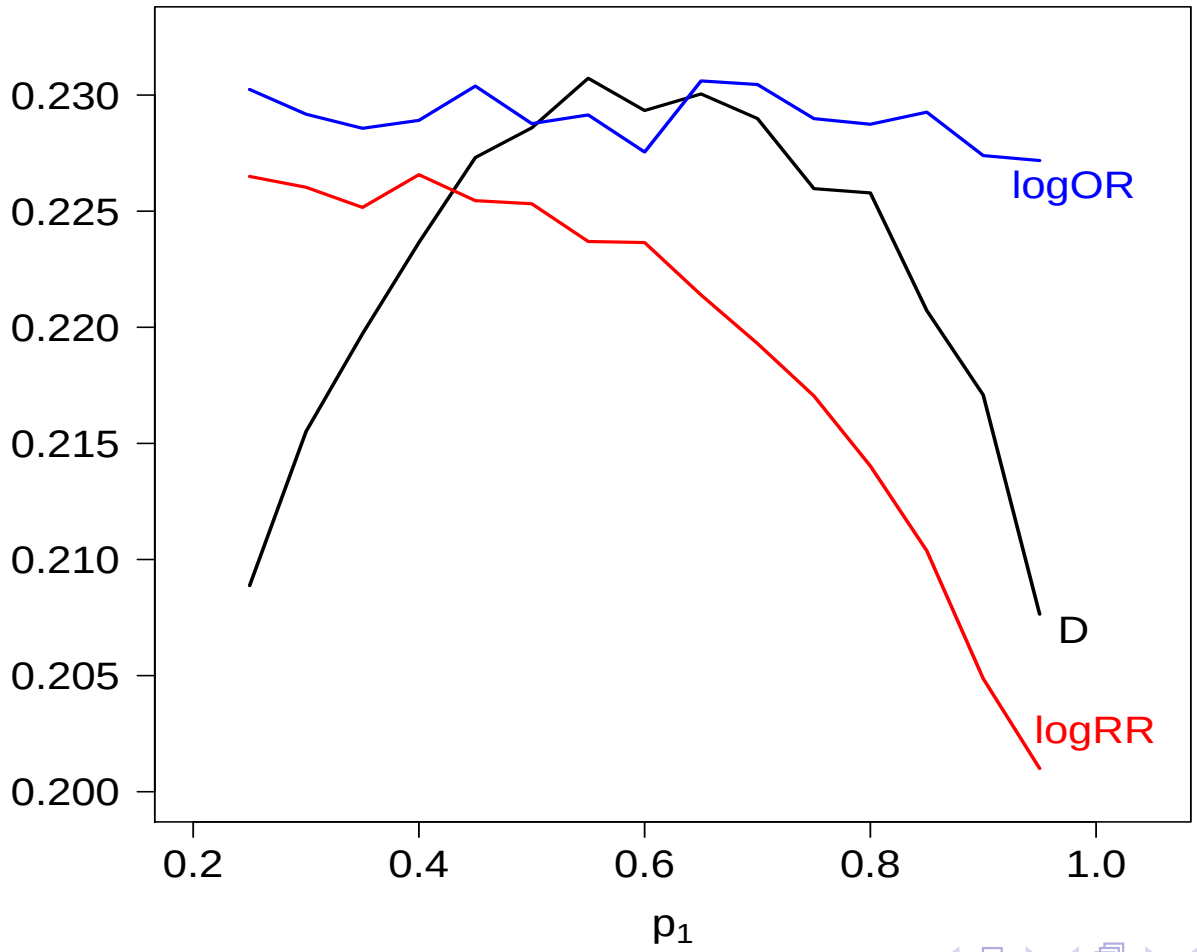
Average Length of the Confidence Interval

- $D = \log RR = \log OR = 0.2$ fixed, $p_1 = 0.25, 0.3, ..., 0.95$



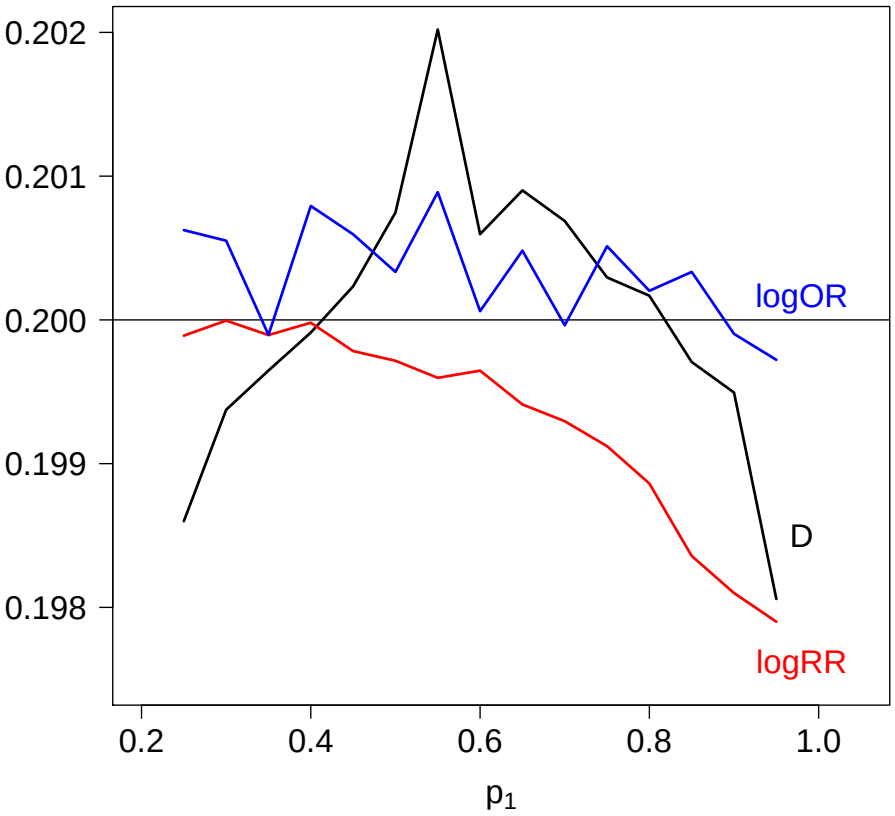
Median unbiased Estimation

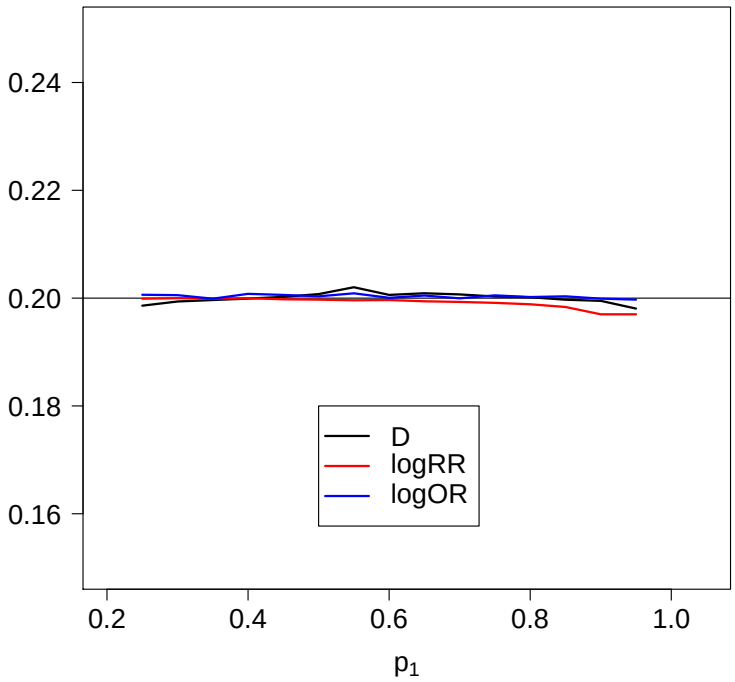
- $D = \log RR = \log OR = 0.2$ fixed, $p_1 = 0.25, 0.3, ..., 0.95$



Empirical Median of the Estimator

- $D = \log RR = \log OR = 0.2$ fixed, $p_1 = 0.25, 0.3, ..., 0.95$





Final Remarks

risk difference:

- using test statistics with an improved estimator of the variance of $\hat{D}$ or continuity corrected test statistics results in less liberal confidence intervals

  (Stansen and Hartung, talk at the conference "Evaluation im Gesundheitswesen", Bochum, 2006)

logarithmic risk ratio:

- improvements for the estimator of the variance of $\log \widehat{RR}$ have been worked out

  but: $\log RR$ is not symmetric around $1/2$

  $\Longrightarrow$ improvement is possible on one side only

Final Remarks

logarithmic odds ratio:

- the estimator of the variance of $\log \widehat{OR}$ can be improved by using the results of Hartung and Knapp (2004)

  $\Longrightarrow$ resulting confidence interval is less conservative

References

Fisher, L. (1998). Self–designing clinical trials. *Statistics in Medicine*, **17**, 1551–1562.

Hartung, J. (2001). A self-designing rule for clinical trials with arbitrary response variables. *Controlled Clinical Trials*, **22**, 111–116.

Hartung, J. (2006). Flexible designs by adaptive plans of generalized Pocock– and O'Brien–Fleming–type and by self–designing clinical trials. *Biometrical Journal*, **48**, 521–536.

Hartung, J., Knapp, G. (2004). Improved tests of homogeneity in randomized controlled multi–center trials with binary outcome. *Far East Journal of Theoretical Statistics*, **13**, 101-126.

References

Hartung, J., Knapp, G. (2006). Repeated confidence intervals in self–designing clinical trials and switching between noninferiority and superiority. *Biometrical Journal*, **48**, 697–709.

Liu, Q., Chi, G. Y. H. (2001). On sample size and inference for two-stage adaptive designs. *Biometrics*, **57**, 172–177.

Wassmer, G. (2003). Data–driven analysis strategies for proportion studies in adaptive group sequential test designs. *Journal of Biopharmaceutical Statistics*, **13**, 586–603.