

# Sample size distribution - an overlooked source of spurious association in funnel plots

Gerta Rücker, Guido Schwarzer, James R Carpenter

Institute of Medical Biometry and Medical Informatics, Freiburg, Germany

London School of Hygiene and Tropical Medicine, London, UK

`ruecker@imbi.uni-freiburg.de`

Funded by Deutsche Forschungsgemeinschaft  
DFG Research Unit (Forschergruppe 534)

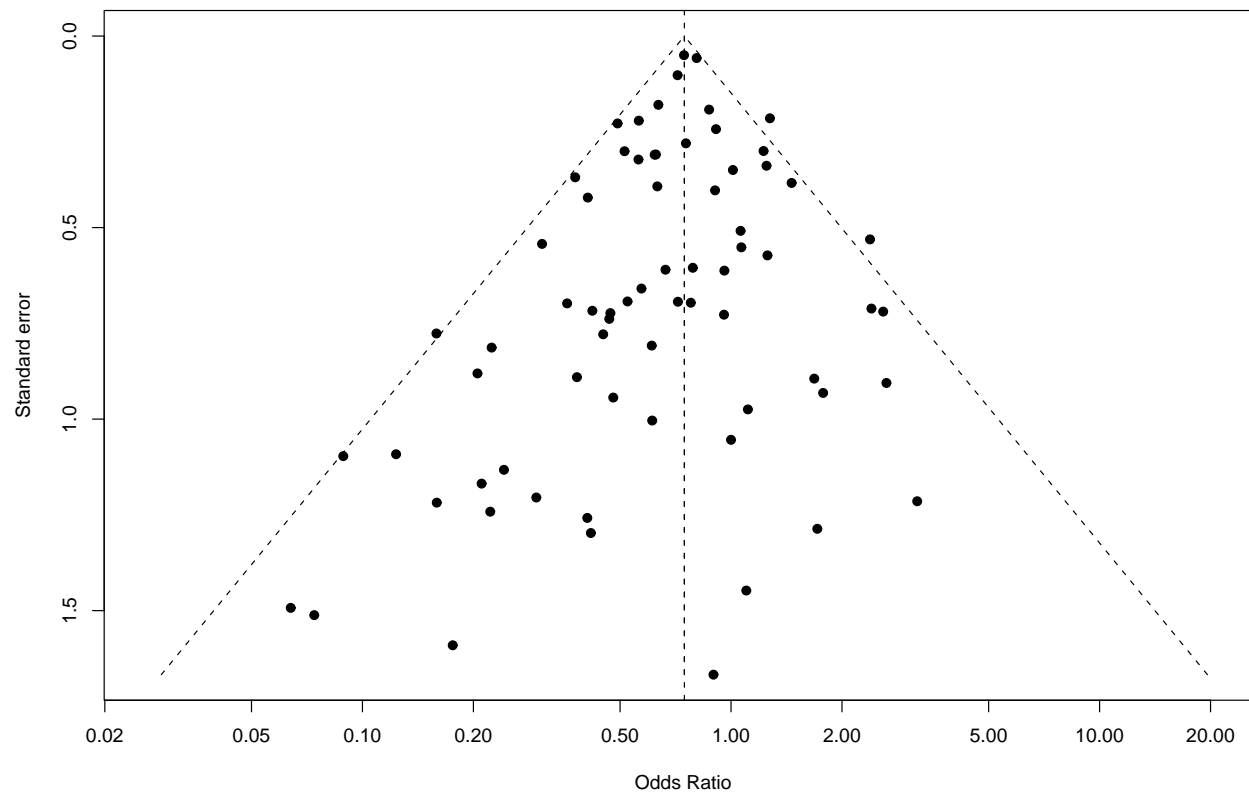
ROeS Seminar 2007, Bern

# Outline

- Background: Recent developments in tests for small study effects (Modified Egger tests, modified rank test, arcsine tests)
- Motivation: Simulation studies evaluating tests for small study effects give different results, depending on the minimum study size
- Investigating study size mix
  - A look at study size distributions
  - A close look at the relationship between the log odds ratio, its standard error and event proportions
  - Effects of study size mix on funnel plot shape
- Small studies and zero studies
- Summary and conclusions

# Funnel plot and publication bias

- Funnel plot: Scatterplot of standard error versus treatment effect estimate
- Tests on publication bias look for asymmetry in funnel plot
- Example: Thrombolytic Therapy after Myocardial Infarction (Oikin 1995)



## **‘Small study effects’: Sources of asymmetry in funnel plot**

- Publication bias: Small studies with stronger effect have an increased chance of being published
- Clinical heterogeneity: Small studies may select patients who are more likely to benefit from the intervention
- Selective outcome reporting: If a prespecified outcome does not show a significant result, another outcome, more promising, may be preferred
- For binary outcomes, mathematical: Standard error of  $\log\text{OR}$  ( $\log\text{RR}$ ) depends on the estimated  $\log\text{OR}$  ( $\log\text{RR}$ )

## Problems of interpreting asymmetry in funnel plots

- Tests for publication bias look for asymmetry in funnel plot
- But: Publication bias is only one of several potential sources of asymmetry, summarized under 'Small study effects'
- Furthermore: Standard tests on publication bias (Begg 1994, Egger 1997) tend to exceed the nominal level
- Recent developments: Harbord (2005), Peters (2006), Schwarzer (2007), Rücker (2007)
- Simulation study (Rücker 2007): Compares new arcsine tests to other recent proposals and standard tests

## Basic idea of arcsine tests: Variance stabilisation by arcsine transformation

- $p_T, p_C$  event proportions in treatment and control arms
- $A \sim \text{Bin}(n_T, p_T), C \sim \text{Bin}(n_C, p_C)$ .

- Variance of  $\arcsin \sqrt{\frac{A}{n_T}}$  approximately equals  $\frac{1}{4n_T}$

- Measure of treatment effect: Arcsine difference

$$\Delta = \arcsin \sqrt{\frac{A}{n_T}} - \arcsin \sqrt{\frac{C}{n_C}}$$

with estimated standard error approximately

$$\sqrt{\frac{1}{4n_T} + \frac{1}{4n_C}},$$

uncorrelated to  $\Delta$  if there is no publication bias

# Arcsine Difference $\Delta$ and Risk Difference

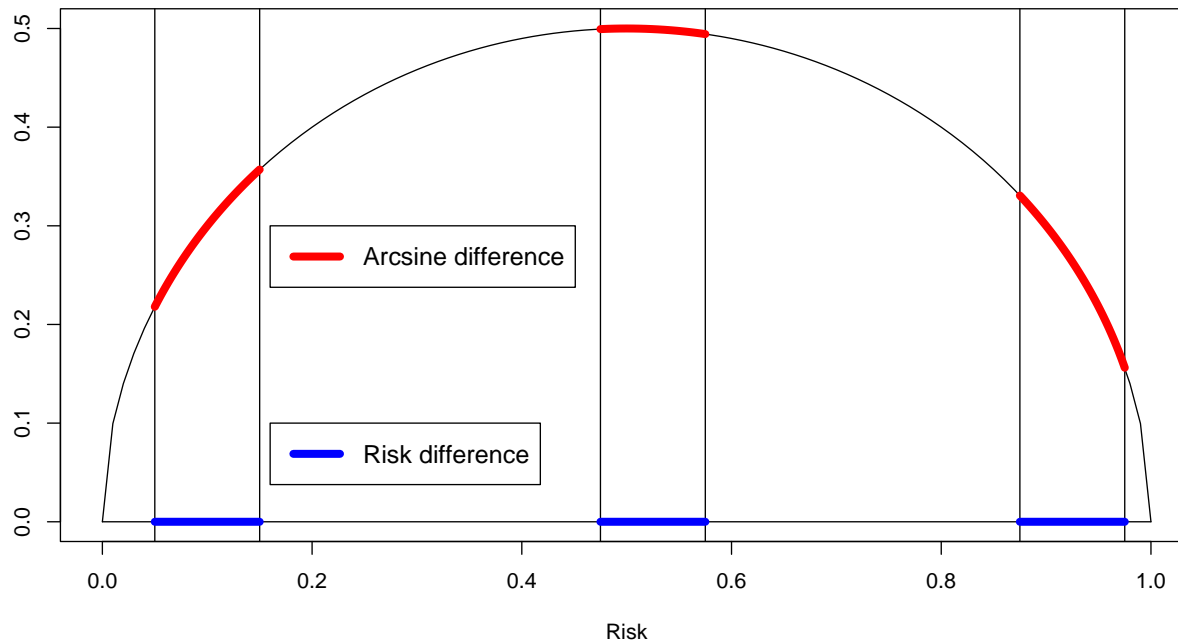
Define function  $f$  of a binomial parameter  $p \in [0, 1]$ :

$$f(p) = \sqrt{p(1-p)}$$

Arcsine difference is the arc length corresponding to the risk difference:

$$\int_{p_1}^{p_2} \sqrt{1 + f'(t)^2} dt = \arcsin \sqrt{p_2} - \arcsin \sqrt{p_1}$$

Graphical representation of the arcsine risk difference as arc length

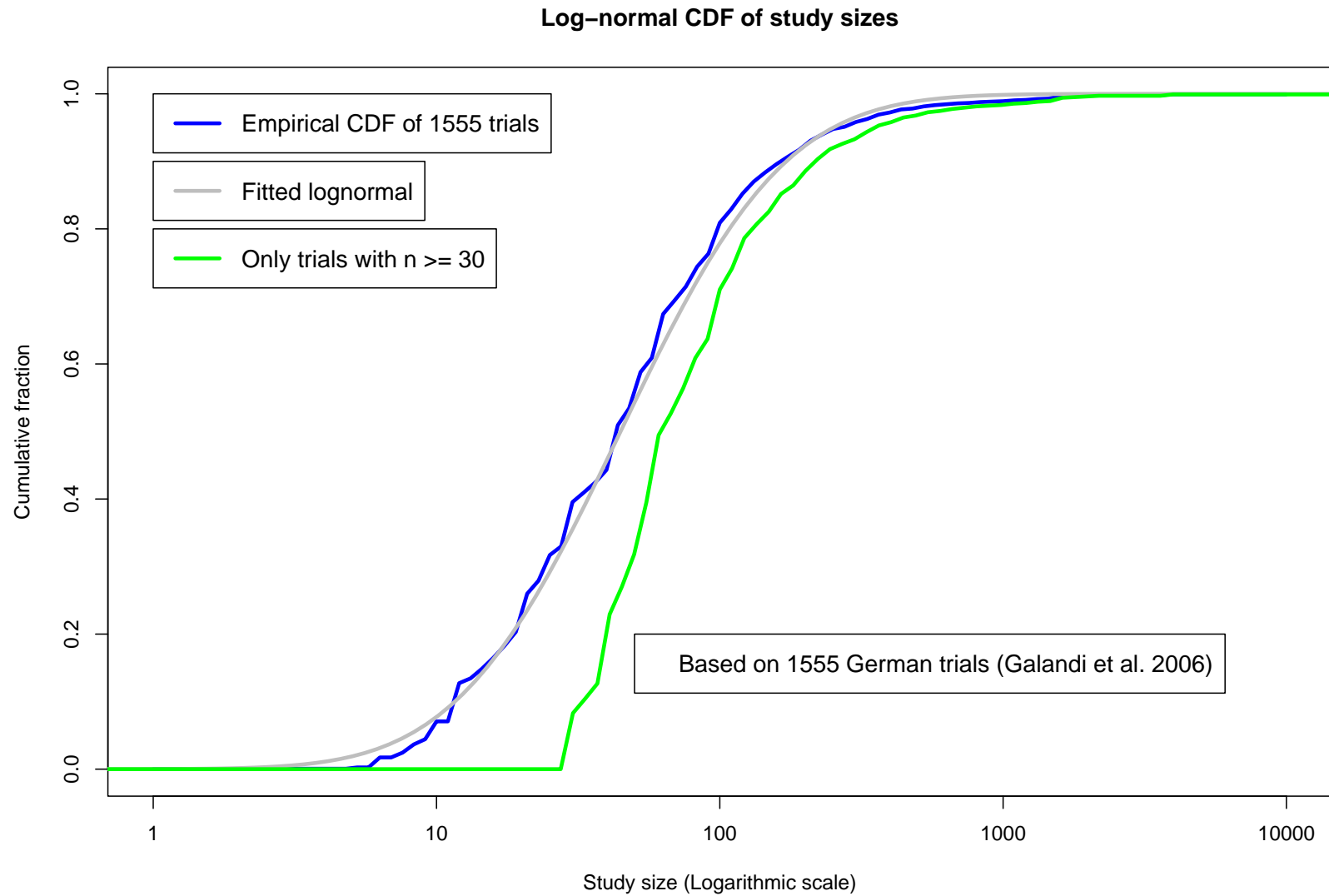


## **Simulation: Source of plausible study size distribution**

Distribution of 1555 trials published in 8 German medical journals  
(Galanti et al. 2006)

- Fit log-normal distribution
- Draw study sizes from this distribution
- Schwarzer (2002, 2007) truncated this distribution by using only trials with  $n \geq 30$

# Study size distribution within 1555 German trials



## Simulation: Study size distribution matters!

- Experience: Type I error (and also power) of tests depend on study size distribution:
  - Excluding small trials ( $n < 30$ ):
    - Egger's test shows inflated type I error
  - Including small trials (trials ranging from 6 to  $> 1000$ ):
    - Egger's test performs no worse than other tests
- A similar occurrence noted before (Macaskill 2001, Harbord 2006) without further investigation
- Aim: Explain these observations!
- Of course, some assumptions about plausible study size distributions unavoidable in simulation studies

## Funnel plot: Standard error in terms of logOR Functional dependence

- Fix study size  $n$  (two groups of equal size) and  $p = p_C$  (true event proportion in control group)
- Calculate usual approximation of standard error SE of  $x = \log \text{OR}$  using  $p$  as parameter
- Result in general ('funnel function')

$$\widehat{\text{SE}}(x) = \frac{2}{\sqrt{np(1-p)}} \sqrt{(1-p)^2 \frac{e^{-x}}{2} + p^2 \frac{e^x}{2} + p(1-p) + \frac{1}{2}}$$

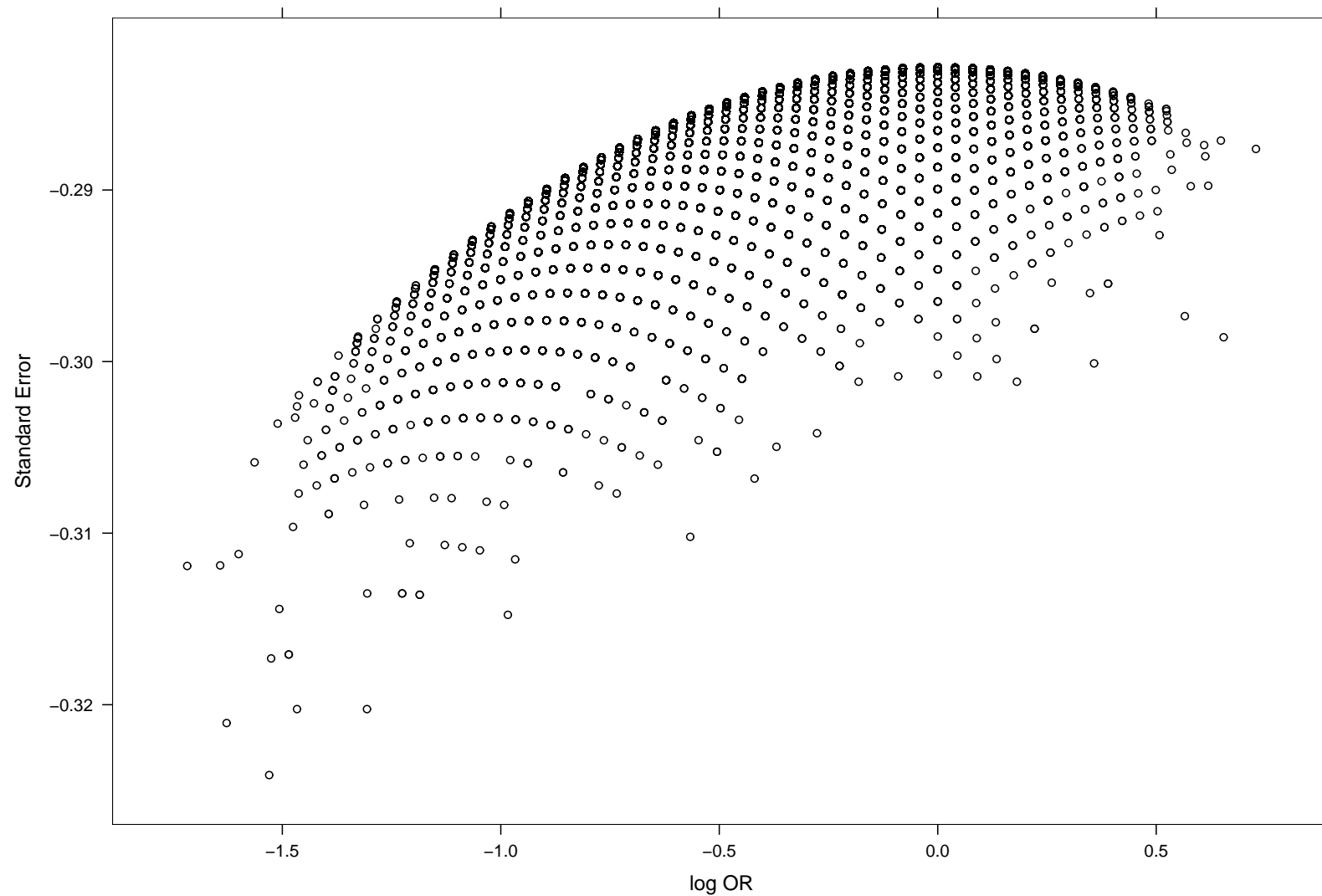
- Special case  $p = 0.5$ :

$$\widehat{\text{SE}}(x) = \frac{2}{\sqrt{n}} \sqrt{\cosh x + 3}$$

- If, instead of  $p_C$ ,  $p = p_T$  is fixed, a similar formula is obtained with  $p$  and  $1-p$  interchanged

# Fixed study size: Funnel plot as a family of catenaries

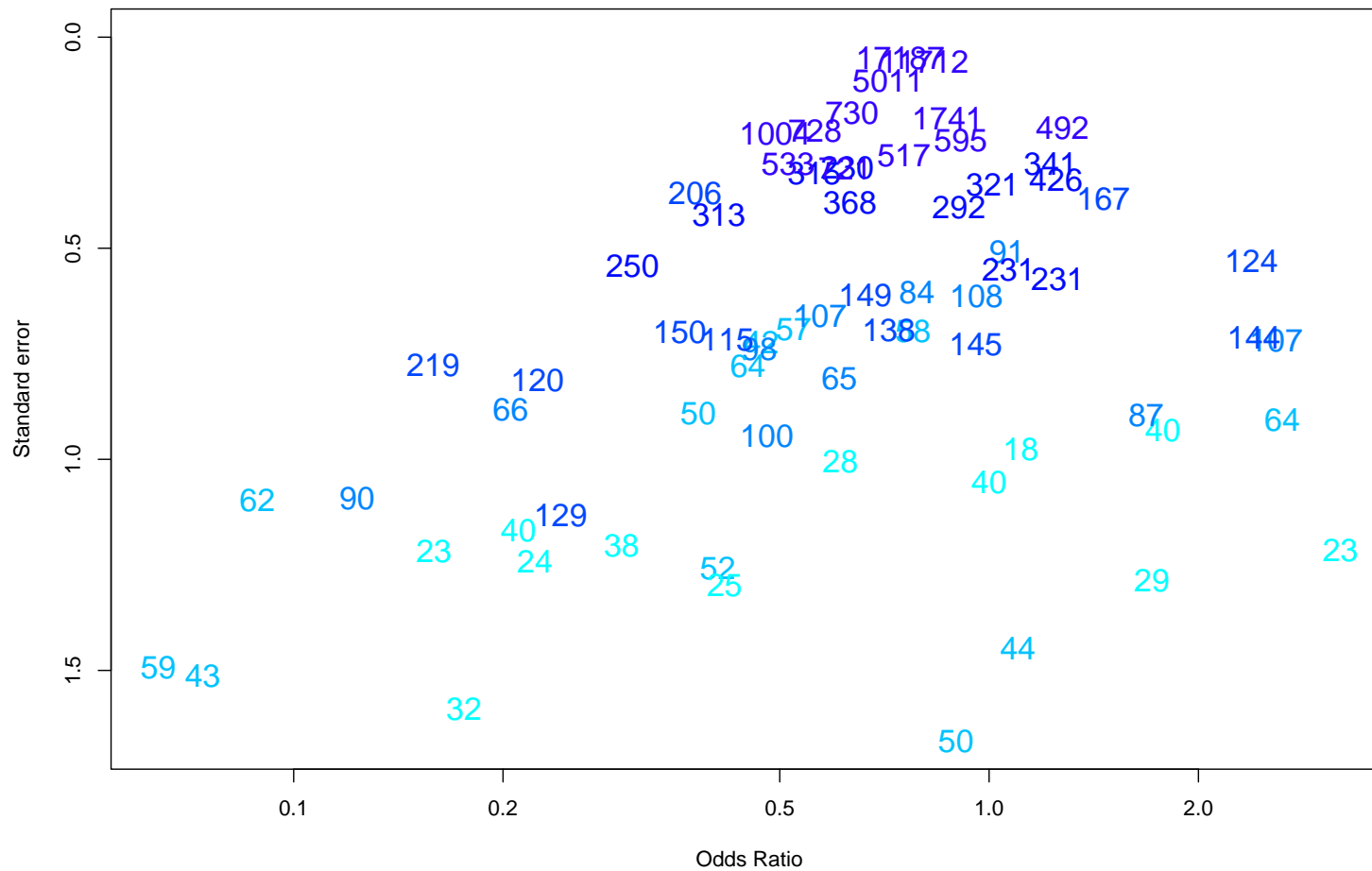
$(n = 200, p_T = 0.4, p_C = 0.5, \text{OR} = 2/3)$



## Study size mix determines funnel plot shape

- For fixed  $n$  (study size) and fixed  $p_C$  (control event proportion), the 'funnel function'  $x \mapsto \text{SE}(x)$  is given by a generalised hyperbolic cosine
- Binomial variation adds noise
- A meta-analysis contains
  - a range of study sizes
  - a range of control event proportions
  - a range of true odds ratios (if heterogeneity exists)
- The funnel plot arises from a mix of these curves
- Its shape depends particularly on the study size distribution (even when there is no heterogeneity!)
- Ignoring small studies may cause bias (see later)

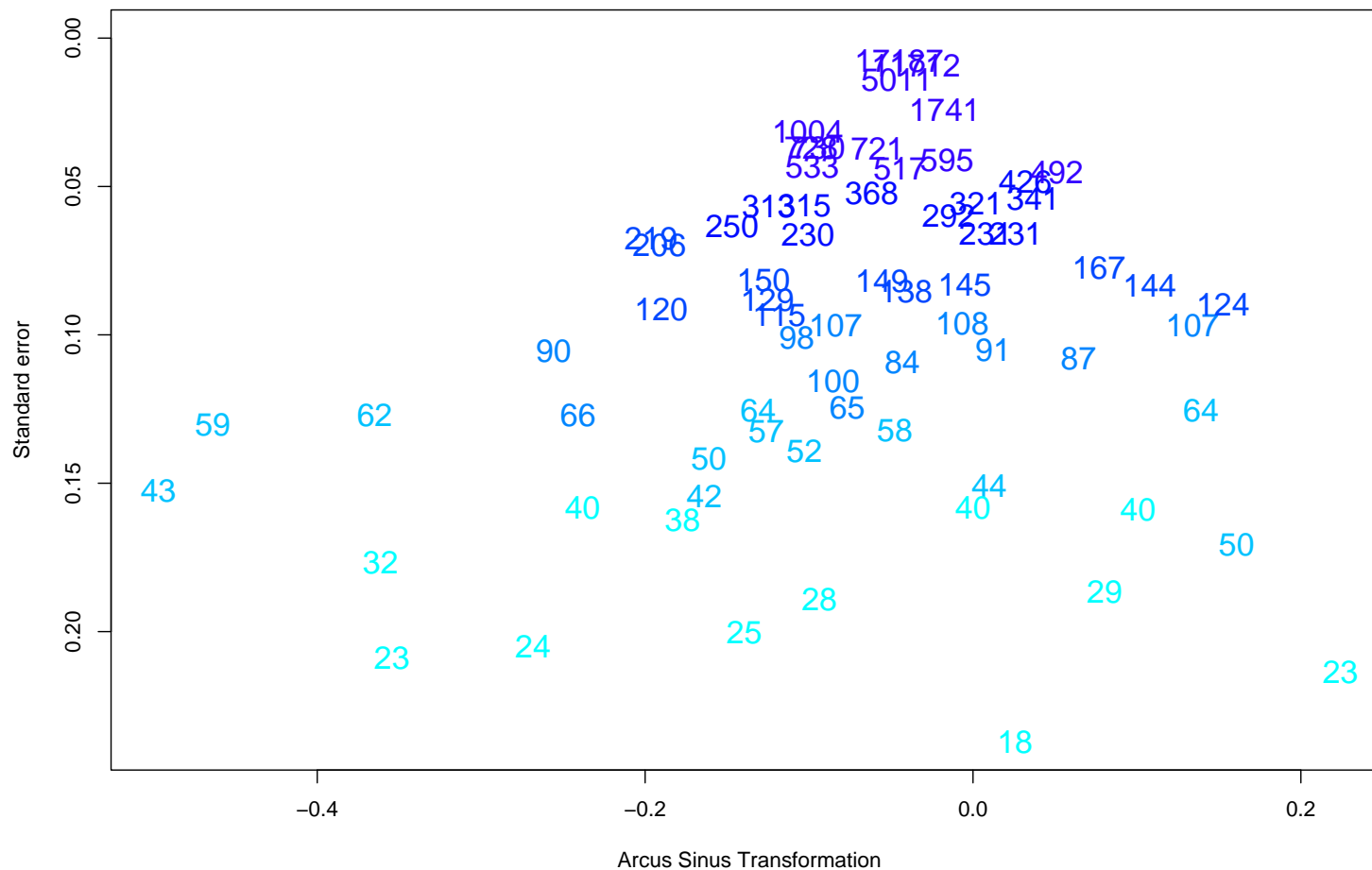
# Study size mix: Funnel plot, grouped by study size Thrombolytic Therapy after Myocardial Infarction (Oikin 1995). Effect measure is logOR



## Study size mix explains our observations

- For large variation of  $n$  (including very small trials) we have
  - a nearly symmetrical plot (if there is no publication bias)
  - not much error inflation in tests for publication bias
- Conversely, if study size varies only little (no small trials) we have to expect
  - spurious asymmetry in the funnel plot
  - spurious publication bias
  - error inflation in tests for publication bias
- For arcsine difference, standard error does not depend on treatment effect:
  - No curving of study size strata in the funnel plot
  - No asymmetry when omitting small studies

# Study size mix: Funnel plot, grouped by study size Thrombolytic Therapy after Myocardial Infarction (Olkin 1995). Effect measure is arcsine difference



## Ignoring or excluding small studies causes bias

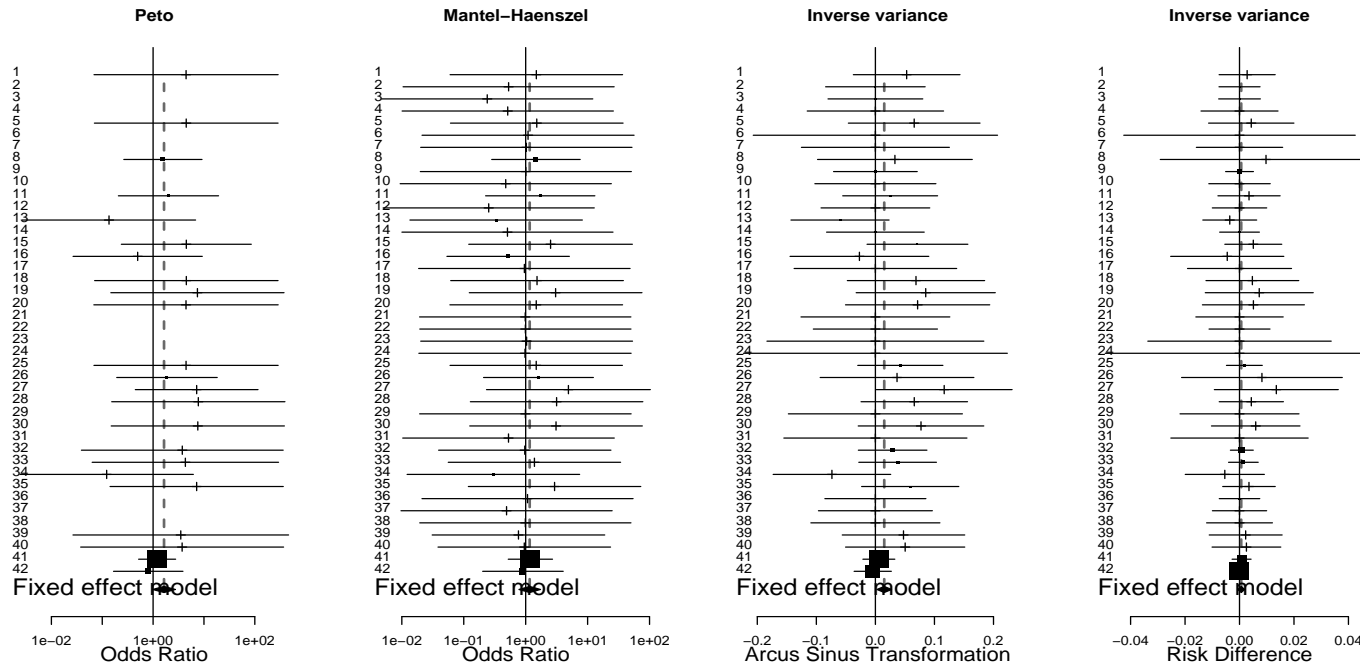
- Particularly: For a meta-analysis with binary outcome with
  - many small studies,
  - small event proportions, and
  - a (maybe large) treatment effect,there will be
  - many 'zero' studies (zero events in each arm)
  - often excluded (because  $\log\text{OR}$  is not defined),
  - thus introducing 'publication bias'!
- Remedies
  - Usual procedure: Add 0.5 (say) to each cell of each table – somewhat arbitrary, can cause reversion of effect direction
  - No need of excluding studies when replacing  $\log\text{OR}$  with arcsine difference

## Ignoring large studies without events can also cause bias

**Example:** Effect of Rosiglitazone on the risk of myocardial infarction and death from cardiovascular diseases (Nissen and Wolski, NEJM 2007)

- Outcome: Death from cardiovascular diseases
- 42 studies including a total of 27833 patients, of whom 61 died (0.22%)
- Groups strongly unbalanced for many studies
- Analysis with Peto's odds ratio:  $\widehat{OR} = 1.64$  [0.980; 2.744] with p-value 0.0596
- But: 19 studies ( $n$  between 77 and 959) ignored because there were no cardiovascular deaths observed

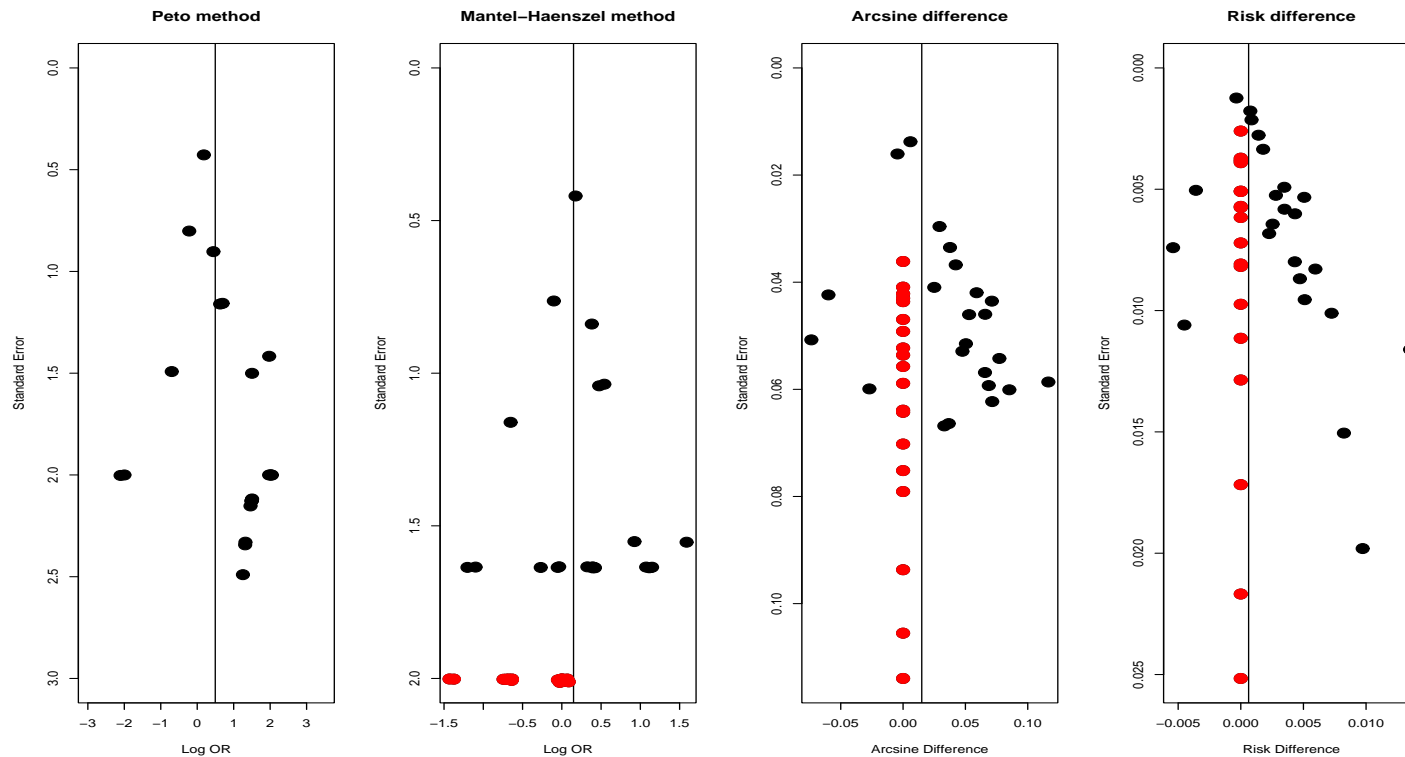
# Rosiglitazone meta-analysis: Forest plots



Do excluded studies provide information about treatment effect?

- 'No' (because there were no events), or rather
- 'Yes', because they indicate that the difference between groups is only small?

# Rosiglitazone meta-analysis: Funnel plots



- Large differences between variance estimators of zero studies
- Large differences between estimated treatment effects
- Large differences between funnel plot shapes

## Rosiglitazone meta-analysis: Results

Outcome: Cardiovascular death

Measure	Method	Number studies included	Increment 0.5 added	Pooled estimate [95% CI] Fixed effect model
Odds ratio	Peto	23	yes	1.64 [0.98–2.74]
Odds ratio	Mantel-Haenszel	23	yes	1.30 [0.83–2.04]
Odds ratio	Mantel-Haenszel	all (42)	yes	1.16 [0.78–1.73]
Risk difference	Inverse variance	all (42)	yes	0.06% [-0.06%–0.19%]
Arcsine difference	Inverse variance	all (42)	no	0.015 [0.003–0.027]

## Summary I: Study size distribution

- Empirical study size distributions investigated
  - compatible with log-normal distribution
  - may include very small studies which should not be neglected in simulations or real meta-analyses
- Analytical relationship between SE of the logOR and estimated logOR itself
  - shows how asymmetry in the funnel plot can result as an artefact of study size mix
- Small studies and sparse events: Excluding studies causes bias
- Zero studies: False conclusions can be avoided by using the arcsine difference

## Summary II: Arcsine transformation

- Variance stabilisation
  - Small asymmetry in funnel plot if there is no bias
  - Reduced error inflation in tests for small study effects
- Properties of the arcsine difference  $\Delta$ 
  - Geometric interpretation as the length of the arc corresponding to the risk difference
  - Often close agreement with logOR, even if there is a range of event proportions (not shown)
- Trials with zero events in each arm need not to be excluded
- Performance of tests for small study effects less dependent on study size distribution

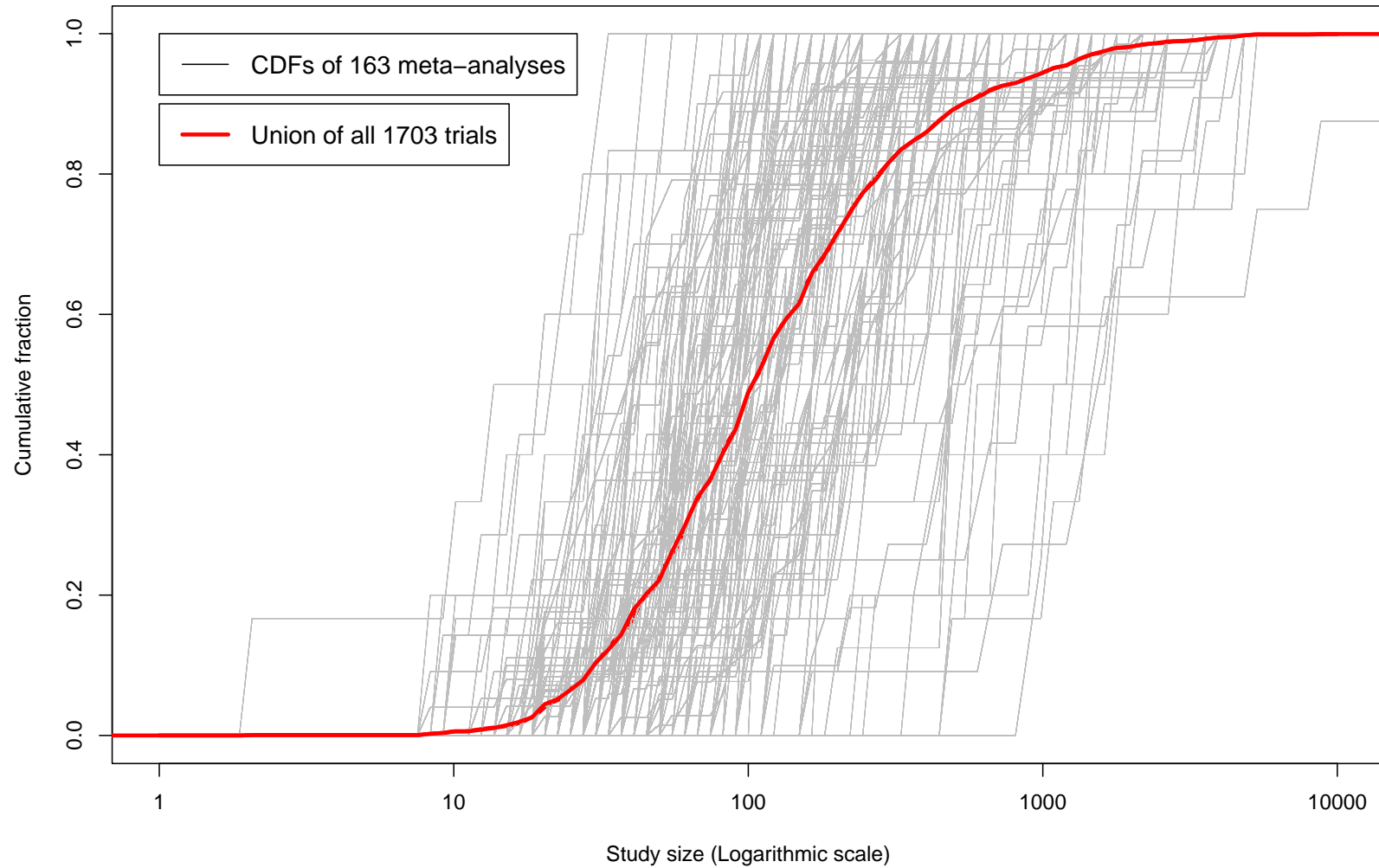
Rücker G, Schwarzer G, Carpenter JR. Arcsine test for publication bias in meta-analyses with binary outcomes. *Statistics in Medicine* 2007; DOI 10.1002/sim.2971.

## Selected References

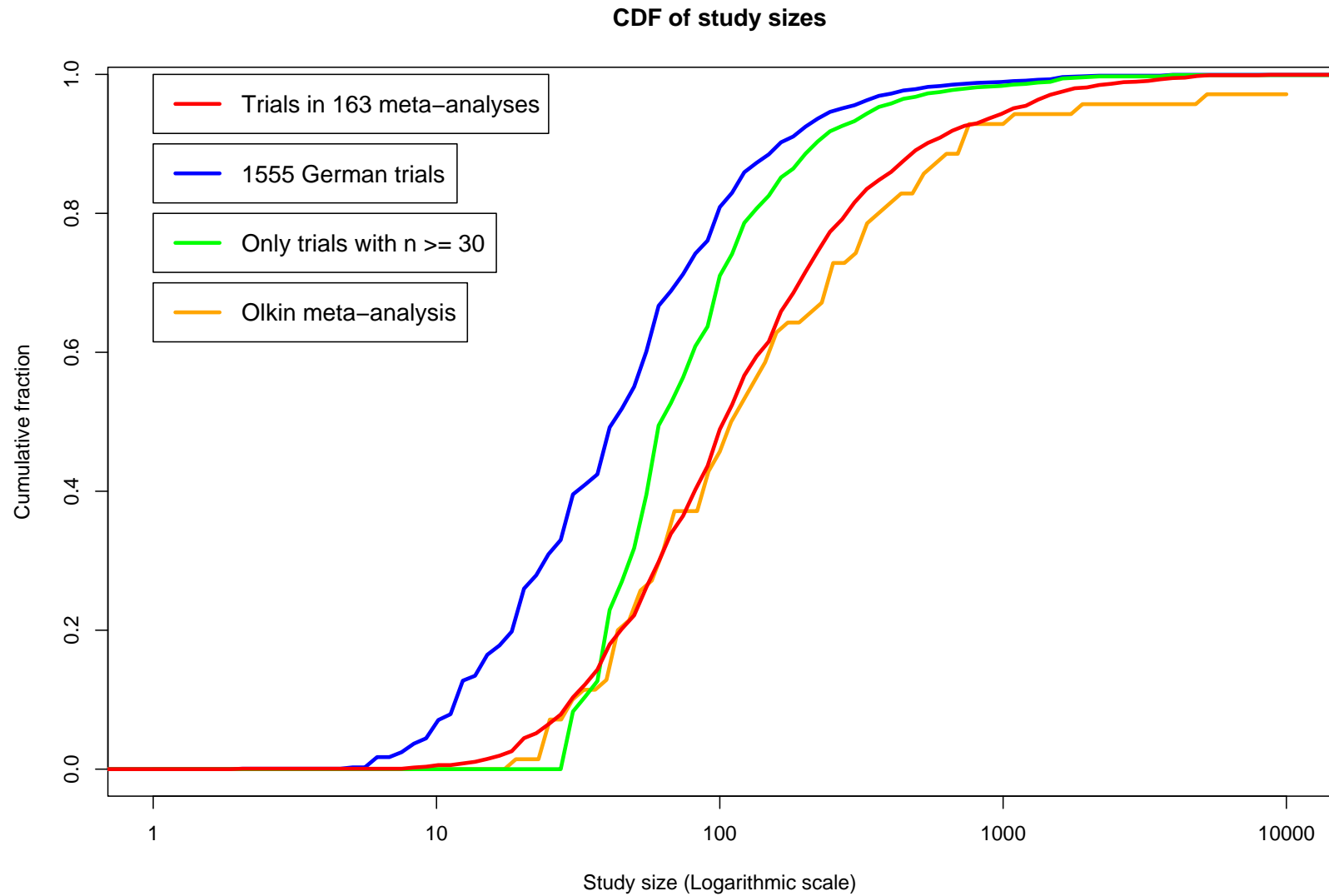
- Olkin I. Statistical and theoretical considerations in meta-analysis. *Journal of Clinical Epidemiology* 1995; **48**:133–146.
- Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal* 1997; **315**:629–634.
- Macaskill P, Walter SD, Irwig L. A comparison of methods to detect publication bias in meta-analysis. *Statistics in Medicine* 2001; **20**:641–654.
- Sweeting MJ, Sutton AJ, Lambert PC. What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of sparse data. *Statistics in Medicine* 2004; **23**:1351–1375.
- Duncan B, Olkin I. Bias of estimates of the number needed to treat. *Statistics in Medicine* 2005; **24**:1837–1848.
- Harbord RM, Egger M, Sterne JA. A modified test for small-study effects in meta-analyses of controlled trials with binary endpoints. *Statistics in Medicine* 2006; **25**(20):3443–3457.
- Peters JL, Sutton AJ, Jones DR, Abrams KR, Rushton L. Comparison of two methods to detect publication bias in meta-analysis. *Journal of the American Medical Association* 2006; **295**:676–680.
- Galanti D, Schwarzer G, Antes G. The demise of the randomised controlled trial: bibliometric study of the german-language health care literature, 1948 to 2004. *BMC Medical Research Methodology* 2006; **6**:30.
- Schwarzer G, Antes G, Schumacher M. A test for publication bias in meta-analysis with sparse binary data. *Statistics in Medicine* 2007; **26**:721–733.
- Bradburn MJ, Deeks JJ, Berlin JA, Localio AR. Much ado about nothing: a comparison of the performance of meta-analytical methods with rare events. *Statistics in Medicine* 2007; **26**:53–77.
- Friedrich JO, Adhikari NK, Beyene J. Inclusion of zero total event trials in meta-analyses maintains analytic consistency and incorporates all available data. *BMC medical research methodology* 2007; **23**:5.
- Nissen SE, Wolski K. Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular diseases. *NEJM* 2007; **356**(24):2457–2471.
- Diamond GA, Bax L, Kaul S. Uncertain effects of rosiglitazone on the risk for myocardial infarction and cardiovascular death. *Annals of Internal Medicine* 2007; Epub ahead of print.

# Appendix: Study size of 1703 trials from 163 meta-analyses

Empirical CDFs of study sizes of trials included in 163 meta-analyses



# Appendix: All study size distributions in one picture



## Appendix: 'Abnormality' of LogOR statistic

- Let  $n_T = n_C = n$ . Let  $a$  be very small and  $c = n - a$ , that is, we assume a very large treatment effect (negative sign).
- The logOR test statistic is

$$\sqrt{\frac{2a(n-a)}{n}} \log \frac{a}{n-a}$$

- For  $a = 1$ , this is of order  $-\log(n-1)$ , but for  $a \rightarrow 0$  it converges to 0
- On the other hand, adding an increment, say 0.5, to each cell entry can cause a reversion of the effect direction!  
Olkin (1995), Trial [44]:

Event	yes	no	Total
Treatment	1	38	39
Control	0	11	11

$$\hat{p}_T > \hat{p}_C, \text{ but } \log \text{OR} = \log\left(\frac{1.5 \cdot 11.5}{38.5 \cdot 0.5}\right) = -0.11$$