

WAS IST IM RAHMEN VON EVIDENZ- BEWERTUNGEN RELEVANT UND WIE WIRD ES BEWERTET?

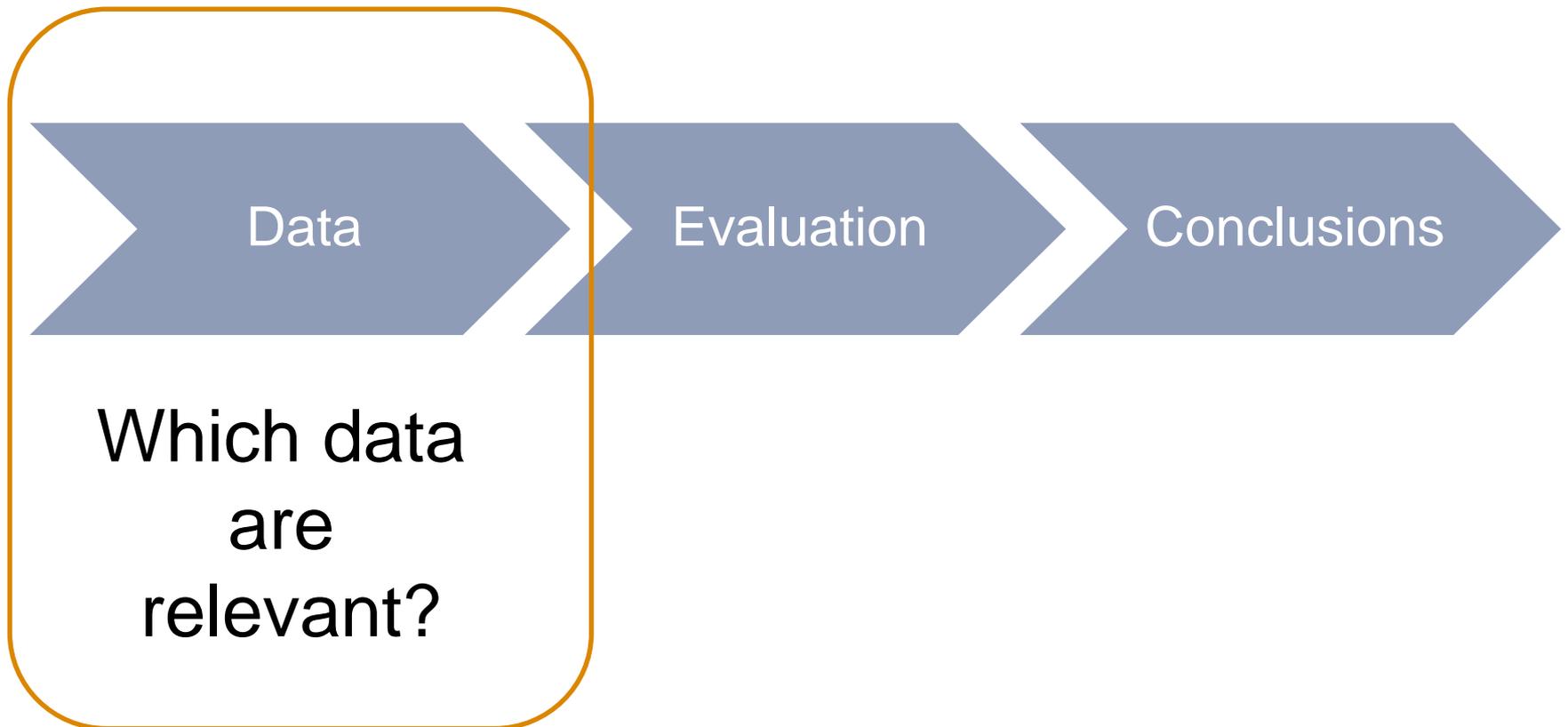
IQWiG im Dialog 2016

Tim Friede

Institut für Medizinische Statistik

Universitätsmedizin Göttingen

GENERATING EVIDENCE



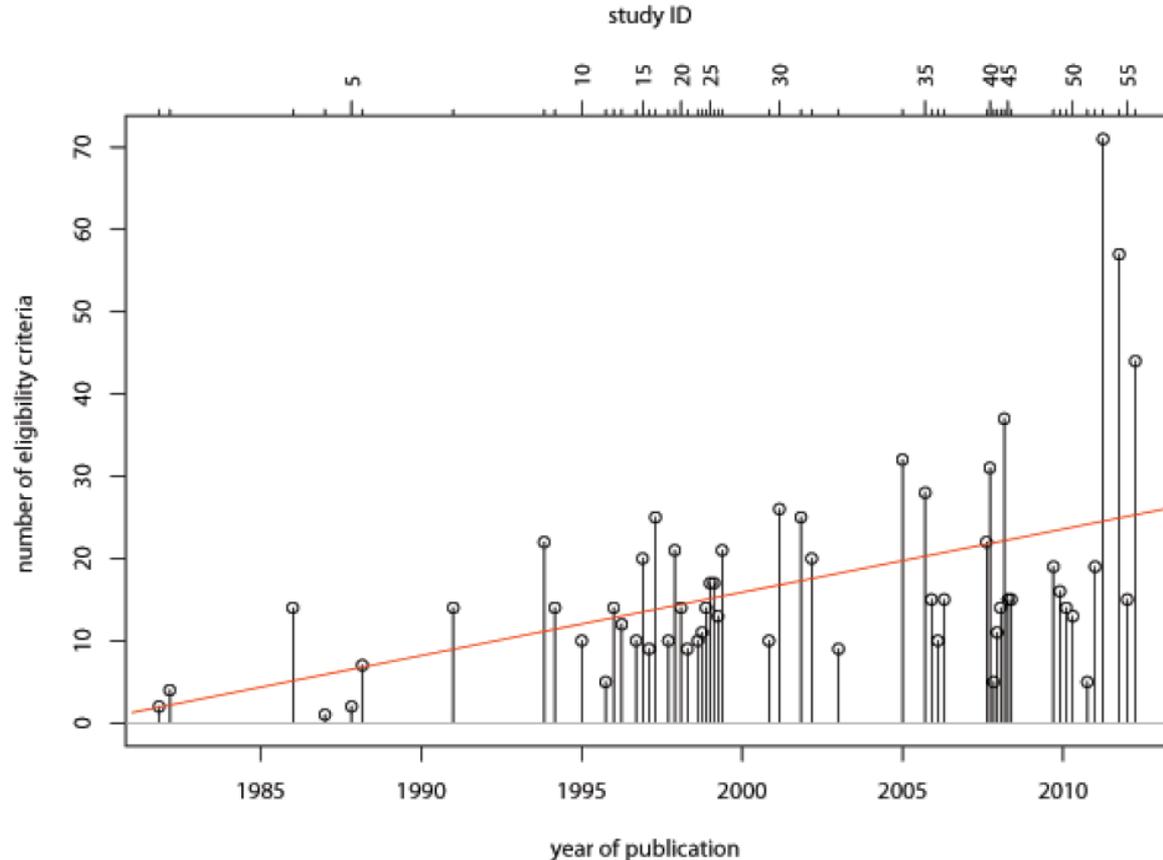
Patient, Population, or Problem	How would I describe a group of patients similar to mine?
Intervention, Prognostic Factor, or Exposure	Which main intervention, prognostic factor, or exposure am I considering?
Comparison or Intervention (if appropriate)	What is the main alternative to compare with the intervention?
Outcome you would like to measure or achieve	What can I hope to accomplish, measure, improve or affect?

<http://hsl.mcmaster.libguides.com/content.php?pid=337527&sid=2763810>

- ▷ Study design: e.g. randomized controlled trial

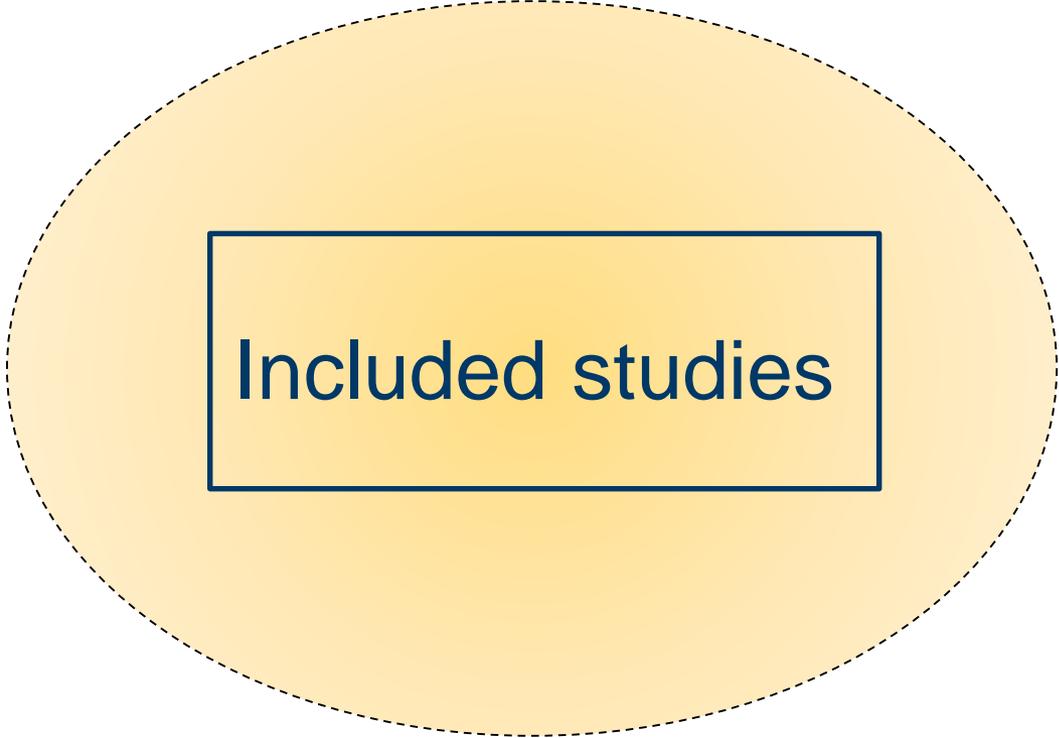
EXAMPLE: RCT ELIGIBILITY CRITERIA BECOMING MORE COMPLEX OVER THE YEARS

- ▶ Example: systematic review of randomized placebo-controlled trials in relapsing multiple sclerosis (Steinvorth et al, 2013)



TO INCLUDE OR NOT TO INCLUDE?

- ▷ PICOS
- ▷ Quality assessment
- ▷ Context



Included studies

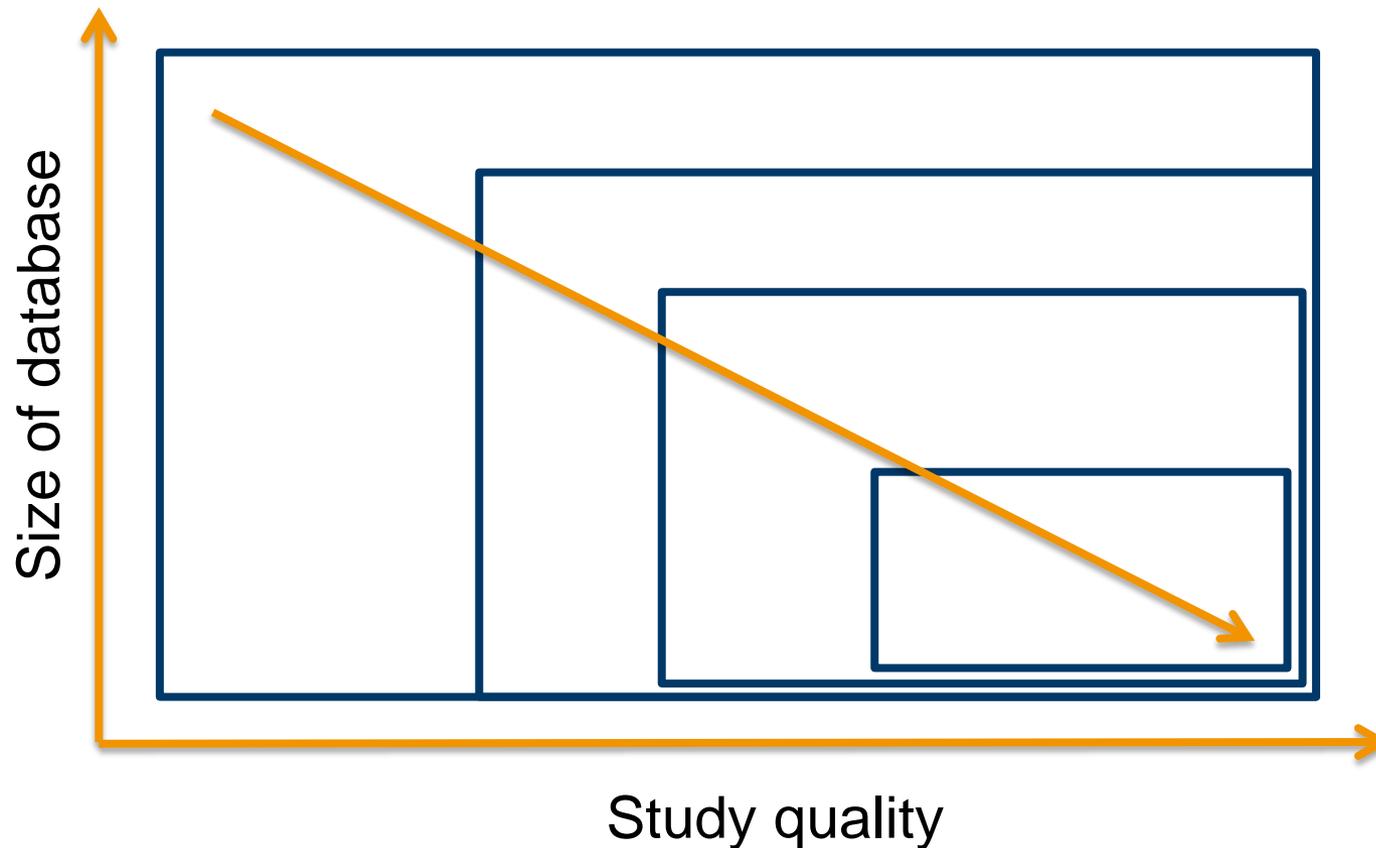
GOOD REASONS FOR RANDOMIZED CONTROLLED TRIALS

- ▶ **Randomised controlled trial:** The gold standard design to evaluate interventions
- ▶ **Contemporary controls** (not historical ones)
- ▶ **Randomisation**
 - ▶ Purpose: Avoid bias due to differences in demographic and clinical characteristics
 - ▶ Principle: known chance to receive each treatment, but not predictable!
- ▶ **Treatment blinding**
 - ▶ Purpose: avoid bias due to differences in treatment or outcome assessment

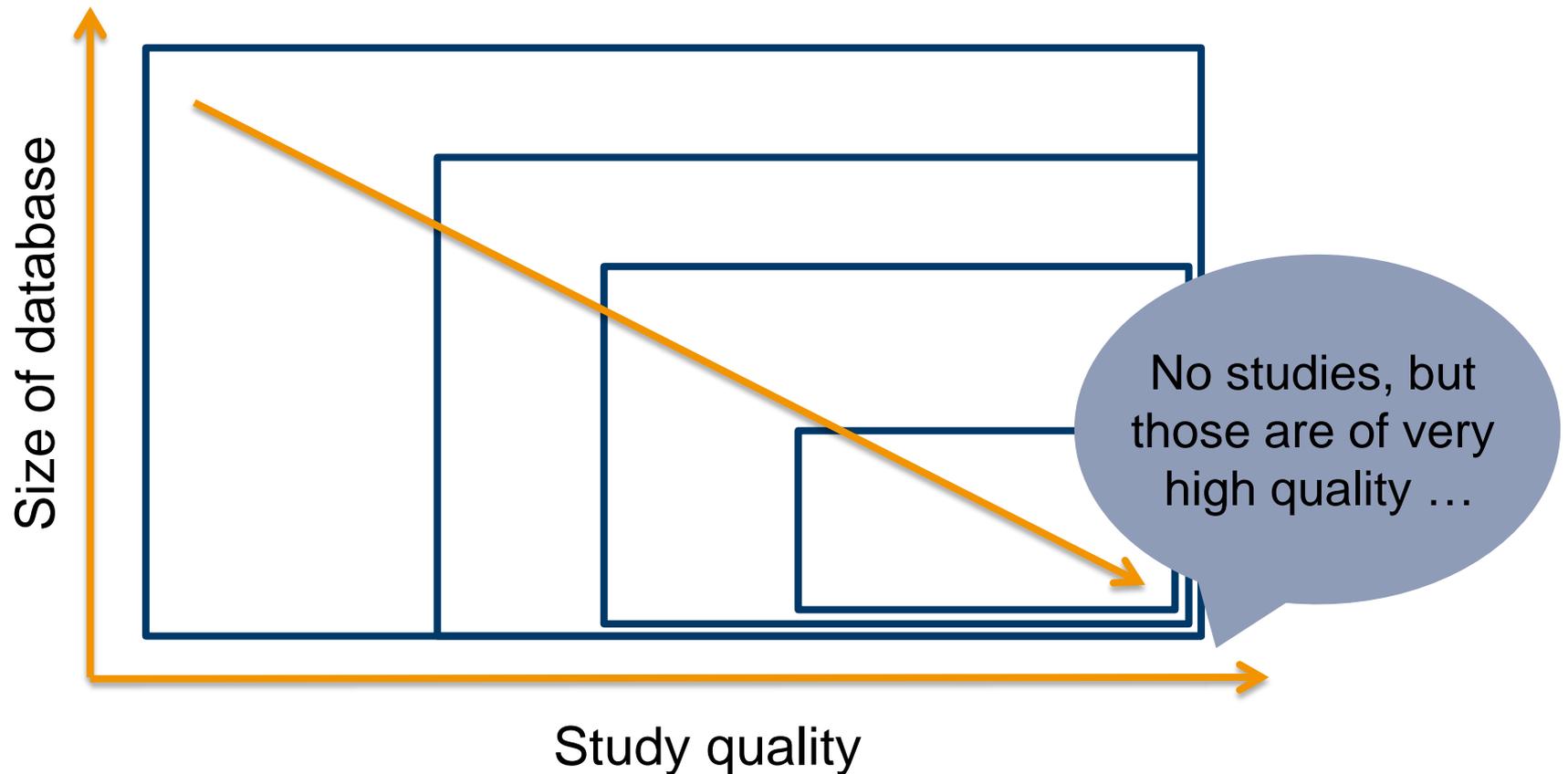
MORE STRINGENT CRITERIA NEEDED?

- ▶ Only prospectively registered trials? (Roberts et al, 2015)
- ▶ How to protect against fraudulent studies?
- ▶ ...

SIZE OF DATABASE VS STUDY QUALITY

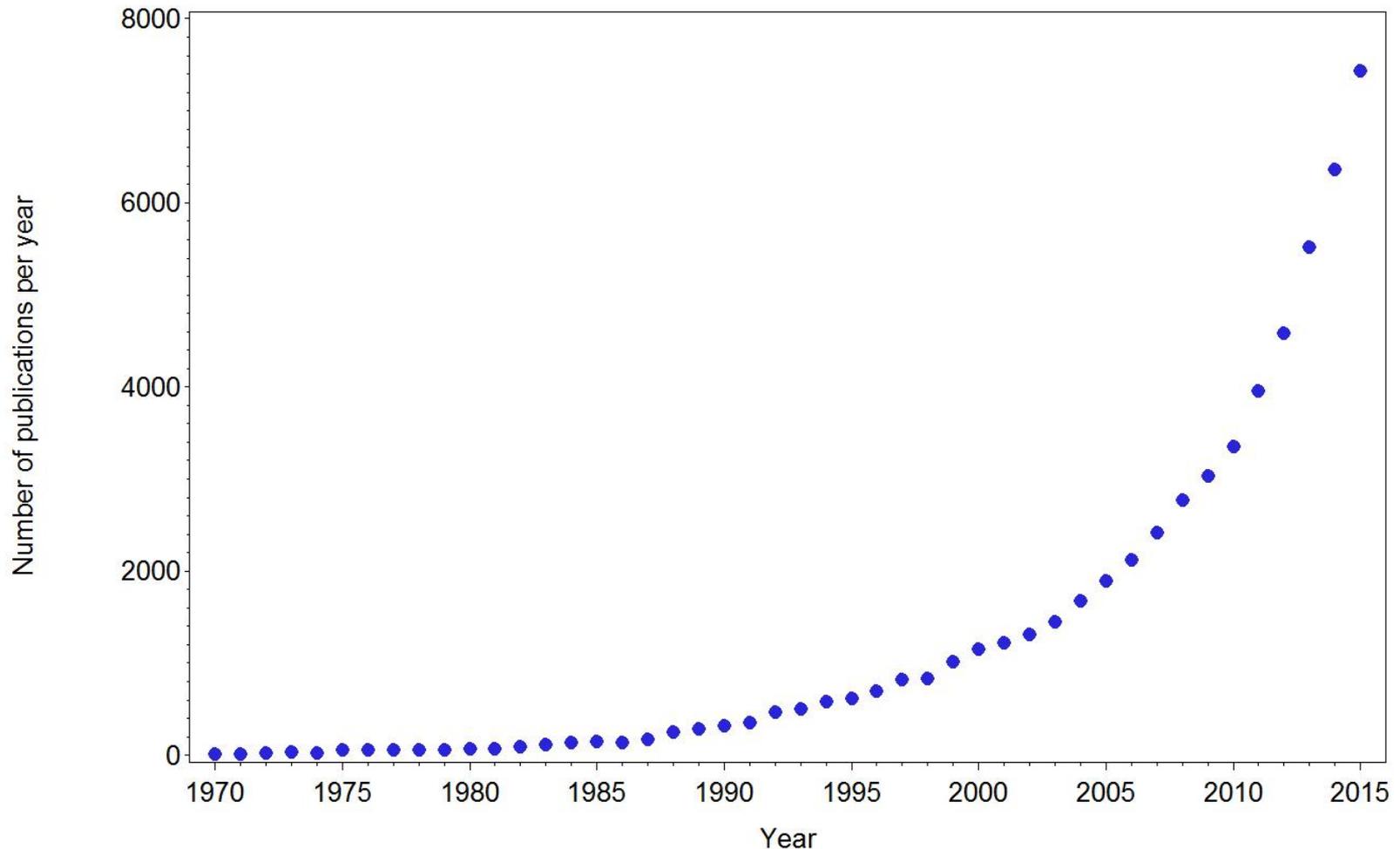


SIZE OF DATABASE VS STUDY QUALITY



GROWING EVIDENCE OUTSIDE RANDOMIZED CONTROLLED TRIALS

▷ **Pubmed search terms: disease registry OR clinical registry**



DEFINITION(S) OF CLINICAL REGISTRIES

- ▶ **Not one, but many definitions in use**
- ▶ Also called **patient registries, clinical data registries, disease registries, outcomes registries, ...**
- ▶ “... a file of documents containing uniform information about individual persons, collected in a systematic and comprehensive way, in order to serve a predetermined purpose.” (Brooke, 1974)
- ▶ “... an organized system for the collection, storage, retrieval, analysis, and dissemination of information on individual persons who have either a particular disease, a condition (e.g., a risk factor) that predisposes [them] to the occurrence of a health related event, or prior exposure to substances (or circumstances) known or suspected to cause adverse health effects.” (US National Committee on Vital and Health Statistics)

DEFINITION(S) OF CLINICAL REGISTRIES



WIKIPEDIA
The Free Encyclopedia

[Main page](#)
[Contents](#)
[Featured content](#)
[Current events](#)
[Random article](#)
[Donate to Wikipedia](#)
[Wikipedia store](#)

Interaction
[Help](#)
[About Wikipedia](#)
[Community portal](#)

[Create account](#)  [Not logged in](#) [Talk](#) [Contributions](#) [Log](#)

Article [Talk](#)

Read

[Edit](#)

[View history](#)

Disease registry

From Wikipedia, the free encyclopedia

Disease or patient registries are collections of secondary data related to patients with a specific diagnosis, condition, or procedure, and they play an important role in [post marketing surveillance](#) of pharmaceuticals.^[1] Registries are different from indexes in that they contain more extensive data.

In its simplest form, a disease registry could consist of a collection of paper cards kept inside "a shoe box" by an individual physician. Most frequently registries vary in sophistication from simple spreadsheets that only can be accessed by a small group of physicians to very complex databases that are accessed online across multiple institutions.^[2]

They can provide health providers (or even patients) with reminders to check certain tests in order to reach certain quality goals.

REQUIREMENTS ON CLINICAL REGISTRY DEPENDING ON PURPOSE

For example ...

- ▶ **Recruitment into RCTs:** only basic information on demographics and disease course required
- ▶ **Epidemiological registry** to estimate prevalence / incidence: capture (nearly) all cases in a certain population
- ▶ Registry **to study natural disease course** / treatment effects: longitudinal data
- ▶ Registry to contribute to **evidence synthesis with randomized controlled trial:** registry needs to be sufficiently similar to RCT in terms of population and endpoints captured

USE OF ROUTINE DATA FOR RESEARCH

- ▶ Routinely collected data (electronic health records)
- ▶ Vision: Use of routine data (electronic health records) in combination with biobanks, imaging, ... (BIG DATA) to develop biomarkers, prognostic / predictive scores, stratify populations ...
- ▶ Current state of affairs: Individual patient data (IPD) meta-analysis of clinical registries or trial data
- ▶ Ethical issues: informed consent? (see Williams and Pigeot (2016) Biom J and editorial by Wegscheider and Friede)

CONSIDERATIONS

▷ **Risk of bias**

- ▷ Focus on higher study / data quality reduces risk of bias

▷ **Precision in estimation**

- ▷ Larger data pools lead to higher precision (unless between-study heterogeneity large)

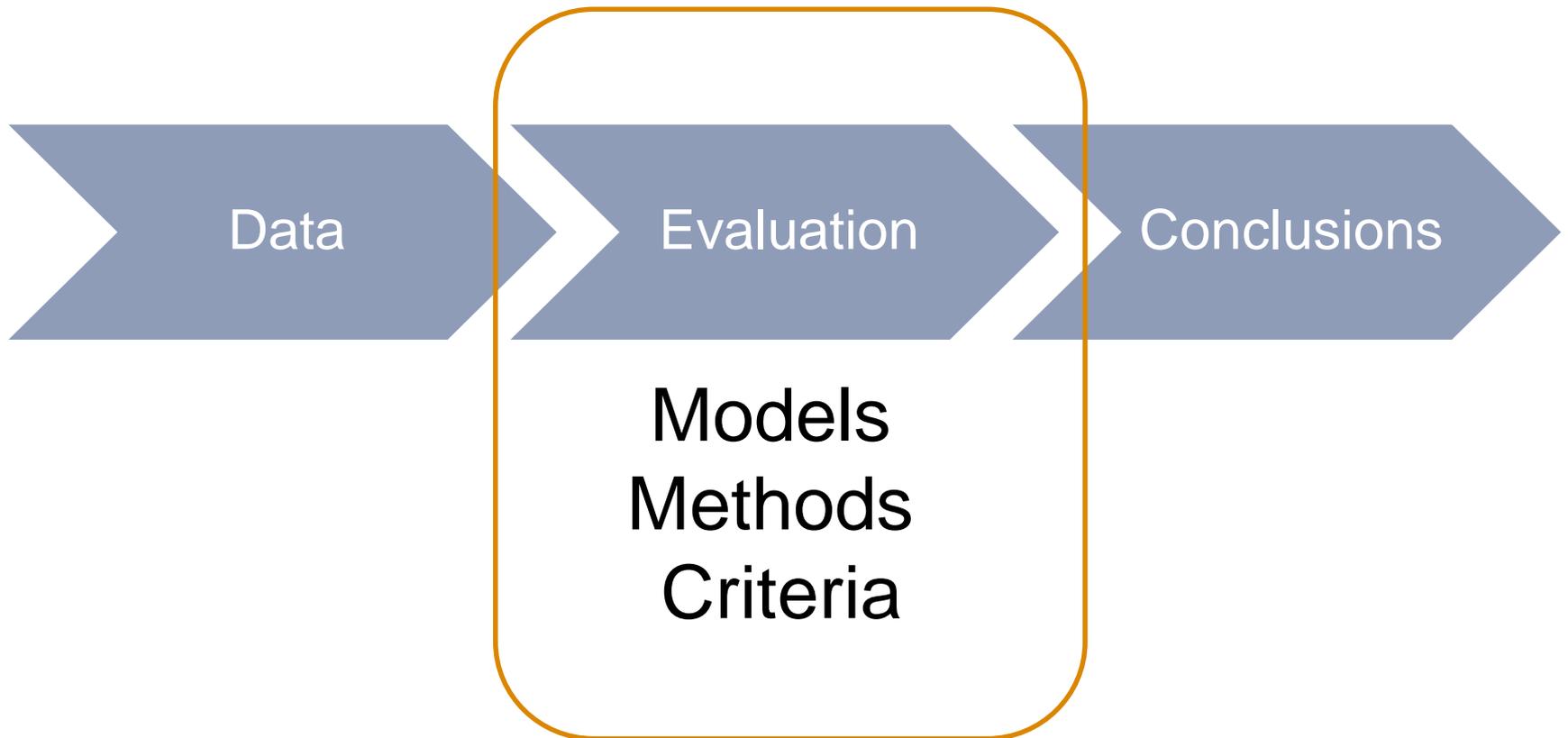
▷ **Resources / time constraints**

- ▷ Reviews focussing on RCTs easier / faster to conduct
- ▷ When including study types other than RCT than eligibility criteria less clear (case series, cohorts, registries, ...)

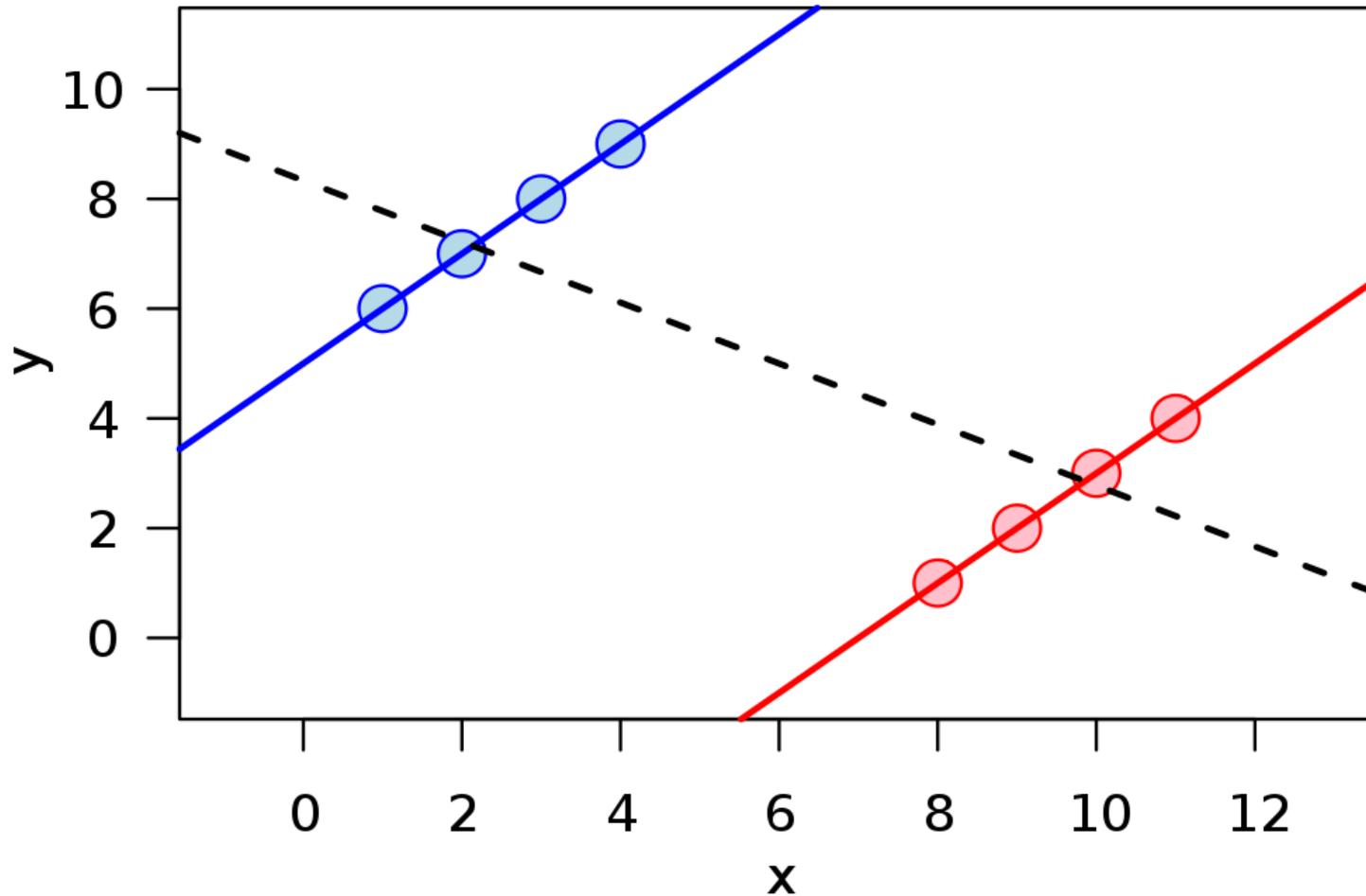
▷ **Interpretation of findings**

- ▷ RCTs often not representative of real life

GENERATING EVIDENCE



SIMPSON'S PARADOX



http://en.wikipedia.org/wiki/Simpson%27s_paradox

META-ANALYSIS

▷ General fixed effect model

- ▷ Assumption: The true (unknown) treatment effects $\theta_1, \dots, \theta_k$ in the studies 1 to k are the same (i.e. $\theta_1 = \dots = \theta_k = \theta$).

- ▷ Treatment effect estimate (with weights w_i):
$$\hat{\theta} = \frac{\sum_{i=1}^k w_i \hat{\theta}_i}{\sum_{i=1}^k w_i}$$

- ▷ Inverse-variance weighted method: $w_i = 1/\sigma_i^2$ (with variance σ_i^2)

- ▷ Formulae for variance depend on effect measure (e.g. mean difference, log-odds ratio)

- ▷ Confidence interval $\hat{\theta} \pm z_{1-\alpha/2} \sqrt{1/(\sum w_i)}$

- ▷ Some **specific methods** for combining odds ratios (e.g. Mantel-Haenszel, Peto)

BETWEEN-STUDY HETEROGENEITY

▷ Fixed effect model

- ▷ assumes no between-study heterogeneity (i.e. $\theta_1 = \dots = \theta_k$)
- ▷ confidence intervals too narrow if heterogeneity present
- ▷ heterogeneity can be assessed graphically or by hypothesis tests (see e.g. Sutton et al (2000), Chapter 3)

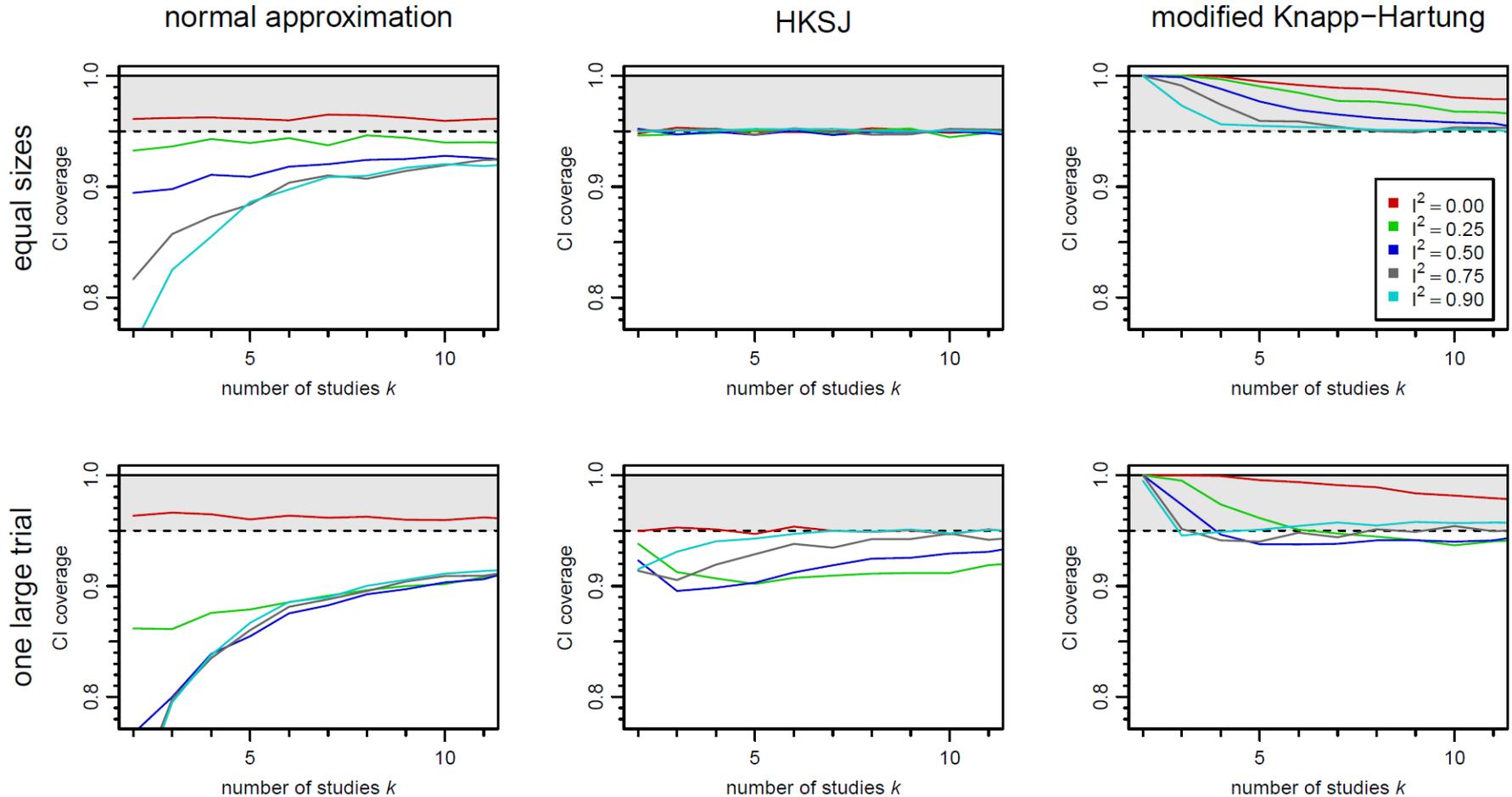
▷ Random effects model

- ▷ Study-specific effect sizes $\theta_1, \dots, \theta_k$ from a normal distribution with mean μ and variance τ^2
- ▷ Hence, the weights become $w_i = 1/(\tau^2 + \sigma_i^2)$
- ▷ Variance of combined effect $\hat{\mu}$: $\hat{\sigma}_\mu = \sqrt{\sum 1/w_i}$
- ▷ See e.g. Chapter 5 in Sutton et al (2000) for an overview

METHODS FOR CONFIDENCE INTERVALS

- ▶ **Normal approximation:** $\hat{\mu} \pm \hat{\sigma}_{\mu} z_{(1-\alpha/2)}$
- ▶ **Hartung-Knapp-Sidik-Jonkman (HKSJ):** $\hat{\mu} \pm \sqrt{q} \hat{\sigma}_{\mu} t_{(k-1);(1-\alpha/2)}$
 - ▶ based on t-quantile with k-1 degrees of freedom
 - ▶ standard error rescaled by factor \sqrt{q} with
$$q = \frac{1}{k-1} \sum_i w_i (\hat{\theta}_i - \hat{\mu})^2$$
- ▶ **Modified Knapp-Hartung (mKH):** $q^* = \max\{1, q\}$

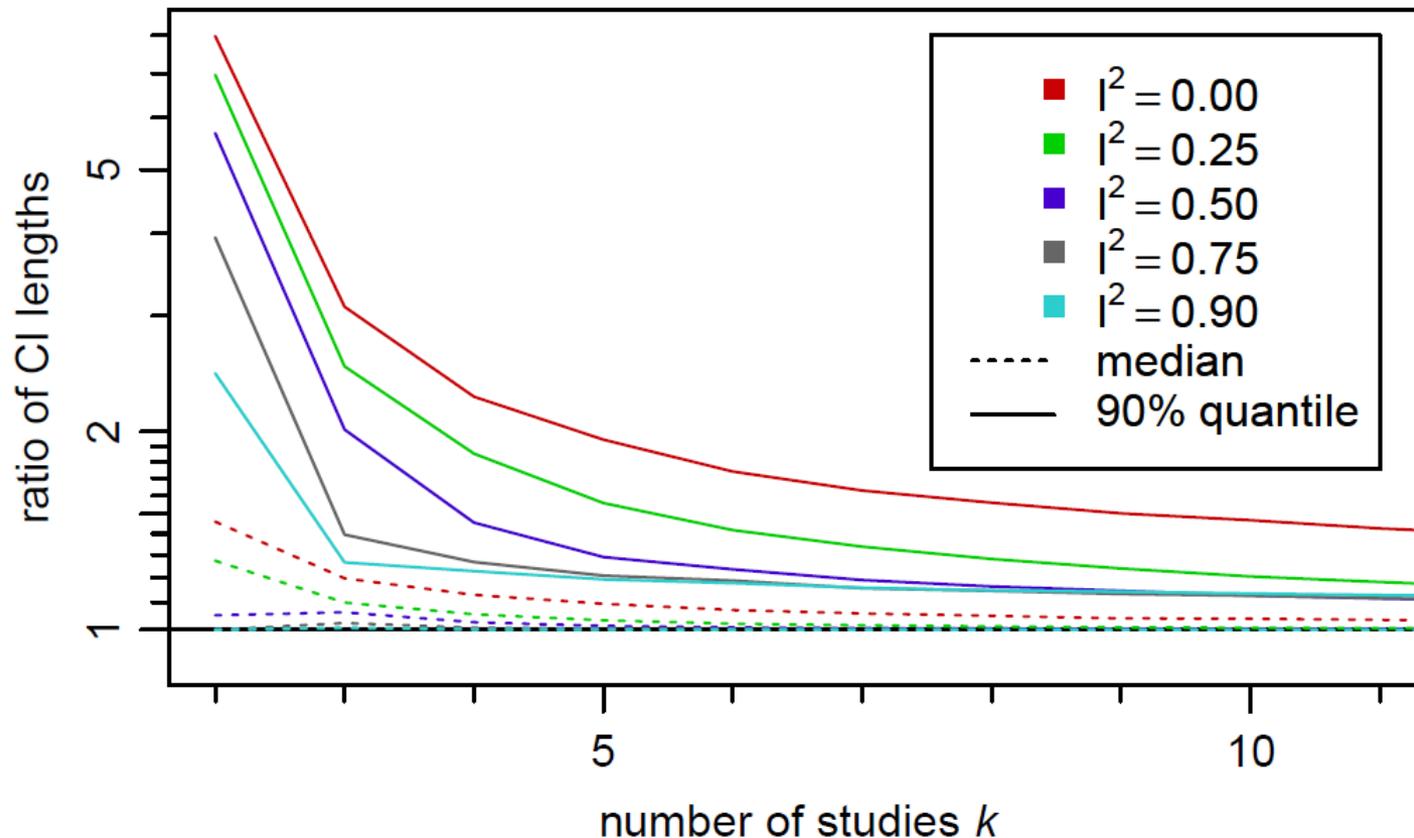
COMPARISON OF METHODS: SIMULATION OF COVERAGE PROBABILITY



IntHout et al, 2014; Röver et al, 2015

SIMULATIONS: INTERVAL LENGTHS

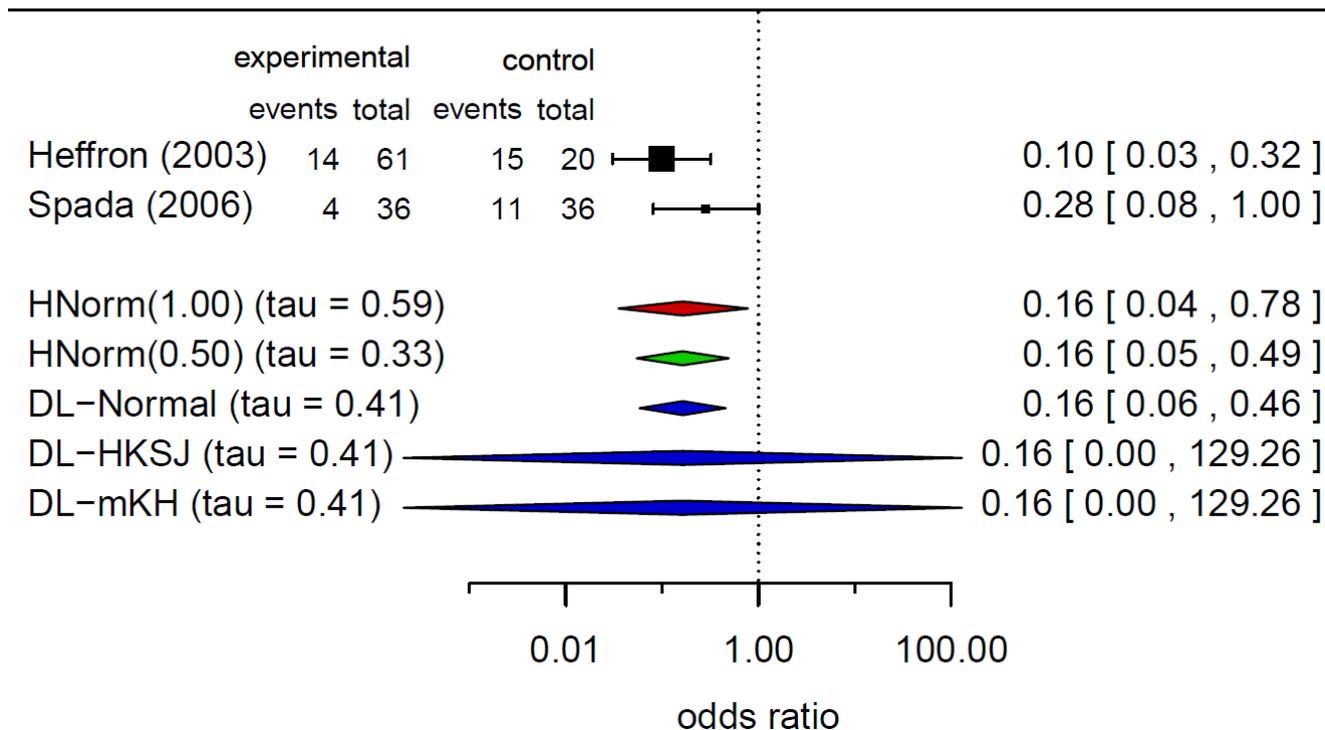
- ▷ Length of mKH confidence interval in relation to length of HKSJ confidence intervals



SPECIAL CASE OF K=2

- ▶ 97.5% quantile of t-distribution with 1 df = 12.7 !!!
- ▶ Examples from Friede et al (2016)

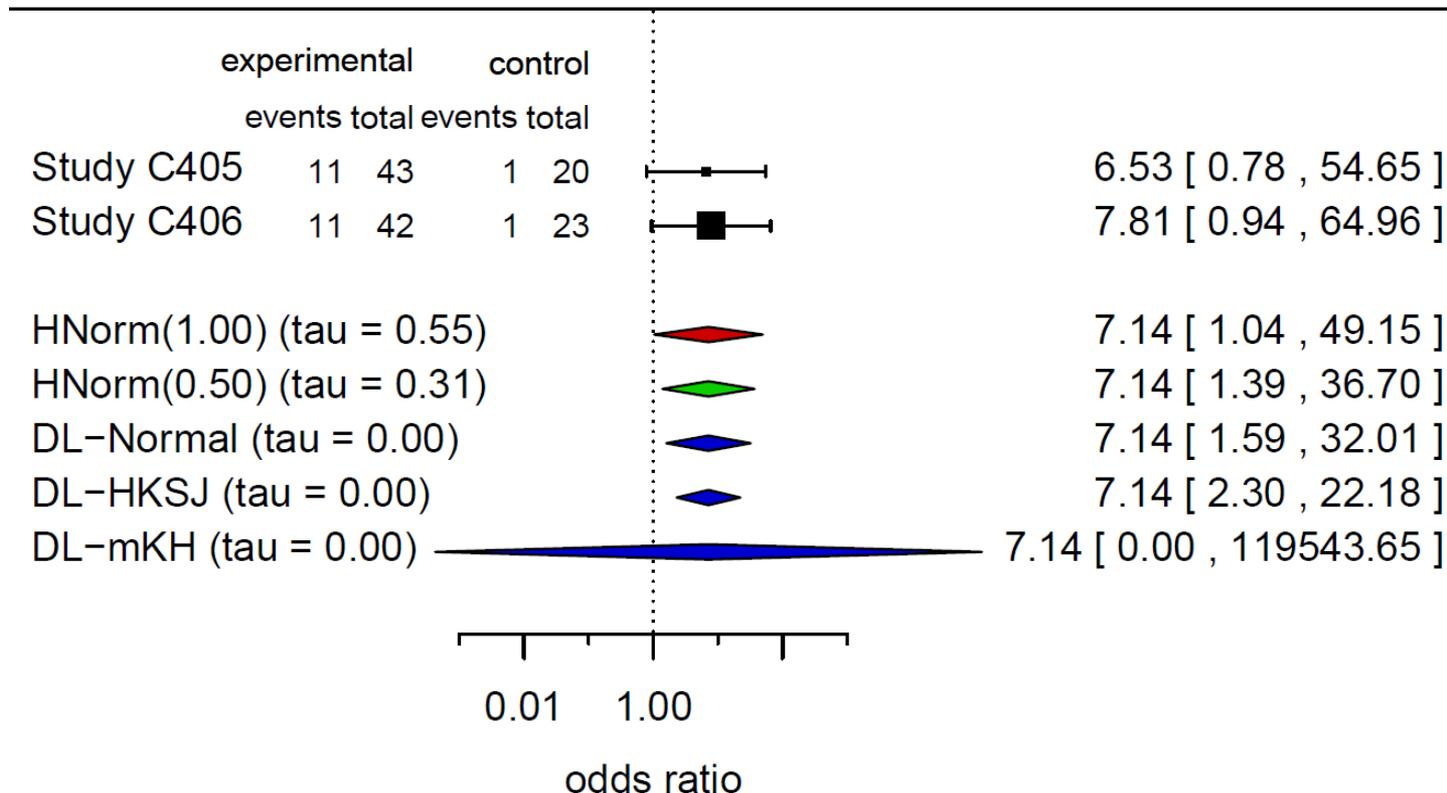
Crins et al. example: acute graft rejection



SPECIAL CASE OF K=2

- ▶ Examples from Friede et al (2016)
- ▶ see also Wiksten et al (2016) **Hartung–Knapp method is not always conservative compared with fixed-effect meta-analysis.** *Statistics in Medicine* 35: 2503–2515.

Krystexxa example: infusion reaction



META-ANALYSES WITH (VERY) FEW STUDIES

- ▶ CI based on **normal** quantiles do not have the right coverage
- ▶ **HKSJ** good coverage if the standard errors similar. In general, however, HKSJ intervals either so wide that they do not allow any conclusion, or very narrow. The latter occurs rarely, but can lead to problematically narrow confidence intervals and unfavourable coverage.
- ▶ **mKH** method yields generally coverage probabilities in excess of nominal level, but intervals generally so wide that they do not allow any meaningful conclusions
- ▶ **Bayesian intervals** appear to be a reasonable compromise (Friede et al, 2016a,b)
 - ▶ R package **bayesmeta** by Christian Röver available from CRAN

SHOULD ESTABLISHED STANDARDS FOR RCT BE APPLIED TO META-ANALYSES?

- ▶ **Registration of protocols**
- ▶ **Adjustment of significance levels**
 - ▶ for multiple endpoints
 - ▶ for sequential testing in cumulative analyses (Trial Sequential Analyses)
- ▶ Sample size calculation
- ▶ **Statistical significance versus clinical relevance**

RANDOMIZED CONTROLLED TRIALS AND SYSTEMATIC REVIEWS / META-ANALYSES

▷ **Randomized controlled trials**

- ▷ Prospective
- ▷ Planned to achieve specific objectives

▷ **Systematic reviews / meta-analyses**

- ▷ Mostly retrospective (exceptions include prespecified meta-analyses of twin studies, preplanned integration of data external to a trial by generalized evidence synthesis)
- ▷ Objectives of the SR / MA might differ from those of the studies included (secondary use of data)
- ▷ At outset often unclear whether sufficiently large datasets can be obtained

REGISTRATION OF PROTOCOLS

▷ **Randomized controlled trials**

- ▷ Clinical trial registries including clinicaltrials.gov, EU Clinical Trials Register, Deutsches Register Klinischer Studien
- ▷ Registration mandatory for certain types of trials
- ▷ Registration expected / required by journals

▷ **Systematic reviews / meta-analysis**

- ▷ Cochrane Collaboration, Campbell Collaboration, PROSPERO
- ▷ No legal requirements; generally, no requirement for publication (exceptions include Cochrane, Campbell)
- ▷ Due to largely retrospective character less important?

HOW IMPORTANT IS THE REGISTRATION OF PROTOCOLS OF SYSTEMATIC REVIEWS?

- ▶ For example: Jacobs et al (2014) argue: “A straightforward way to deal with some of the multiplicity problems is to publish a protocol before the literature search begins ...”
- ▶ Well, this builds on trust that authors do not do the literature search first and then publish the protocol knowing what the outcome will be ...
- ▶ This would be with RCTs generally more difficult / impossible

TRANSPARENCY IN REPORTING OF SYSTEMATIC REVIEWS / META-ANALYSES

- ▶ As with clinical trials, transparency in reporting what was done important for systematic reviews / meta-analyses
- ▶ Reporting guidelines: PRISMA statement for systematic reviews (in analogy to CONSORT statement for RCTs)
- ▶ Some aspects of analyses might be harder to prespecify in SR / MA than RCT (e.g. exploration of between-study heterogeneity). Hence, transparency even more important.
- ▶ Returning to the example above on multiplicity: Transparency is important to allow readers to make their own judgements on significance / relevance of findings (but builds on trust)

MULTIPLICITY DUE TO MULTIPLE ENDPOINTS

- ▶ “The Cochrane Collaboration recommends using up to three primary outcomes – for example, all-cause mortality, serious adverse events, and quality of life [13]. The use of more than one primary outcome (co-primary outcomes) necessitates adjustments of the thresholds for significance because of problems with multiplicity [34].” (Step 3 in Jacobsen et al (2014))
- ▶ **More than one primary endpoint to reflect different dimensions of disease and treatment**
- ▶ **Whether this implies adjustment of the significance level depends on the question asked / hypothesis tested**
 - ▶ At least one endpoint significant? (Union-intersection method)
 - ▶ Does the treatment convince regarding all co-primary endpoints? (Intersection-union method)

MULTIPLICITY DUE TO MULTIPLE ENDPOINTS

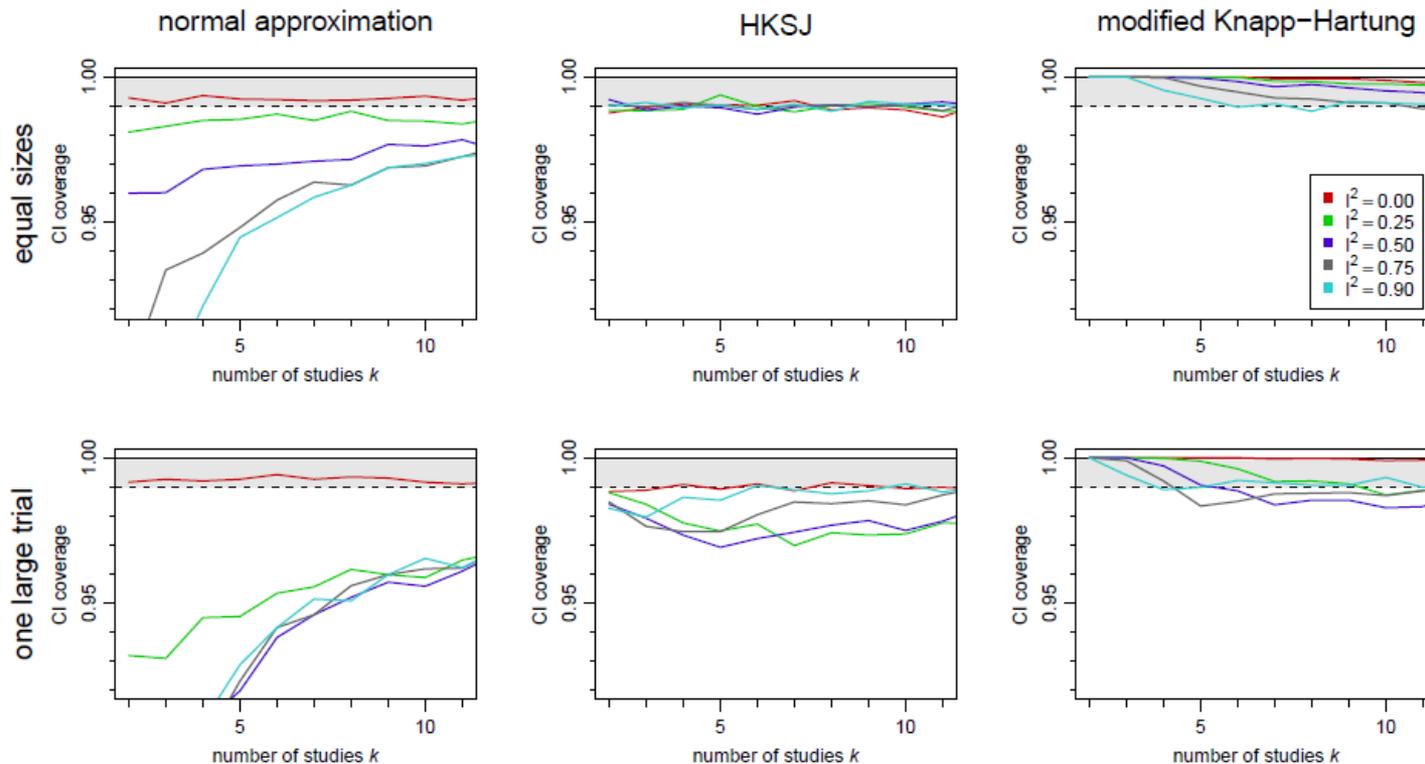
- ▶ **Correlation between endpoints**
 - ▶ usually unknown
 - ▶ can be estimated from individual patient data (if available)
 - ▶ estimation from aggregated data might be possible in some instances (e.g. Friede et al, 2011), but likely with low precision due to small number of studies
- ▶ **Single step procedures** such as Bonferroni procedure: very conservative (in particular with larger number of endpoints and pronounced correlation)
- ▶ **Sequential procedures** (e.g. Bonferroni-Holm): more powerful, but confidence intervals more difficult to construct / uninformative / unavailable

MULTIPLICITY DUE TO MULTIPLE ENDPOINTS

- ▶ “We suggest dividing the pre-specified P-value threshold with the value halfway between 1 (no adjustment) and the number of primary outcome comparisons (Bonferroni adjustment).” (Step 3 in Jacobsen et al (2014))
- ▶ What is the type I error rate of such a procedure depending on the number of endpoints and their correlation?

KEEP IN MIND ...

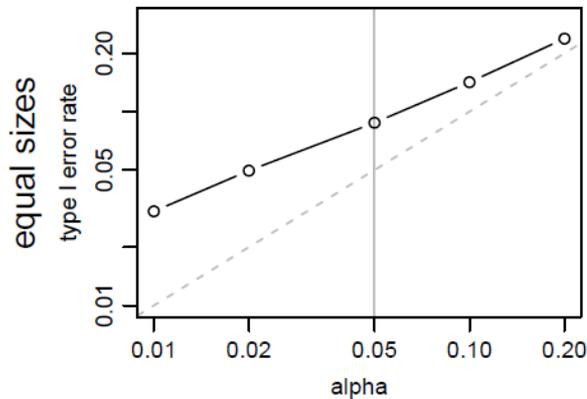
- ▷ ... that we struggle to control the type I error rate when testing one endpoint in meta-analysis with few studies and some degree of heterogeneity
- ▷ This problem does not go away testing at a different level (here 1% instead of 5% above)



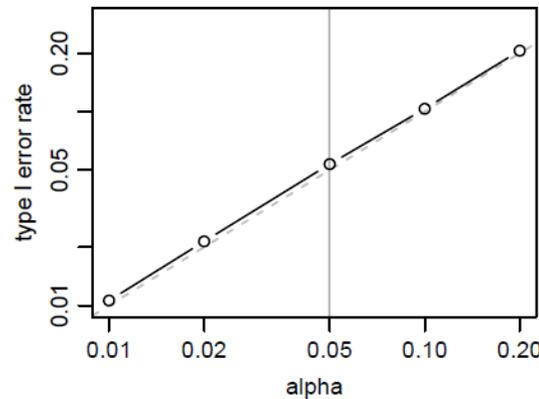
PROBLEM MIGHT EVEN GET LARGER ...

- ▷ Type I error rate vs. nominal significance level
- ▷ Example: $k=5$, $I^2=0.50$

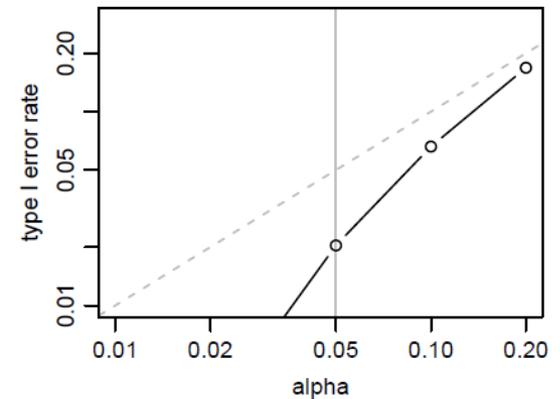
normal approximation



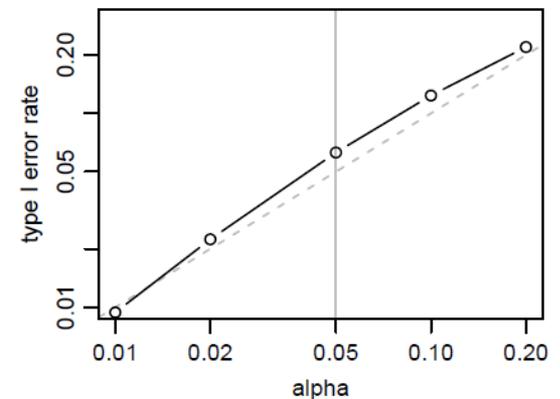
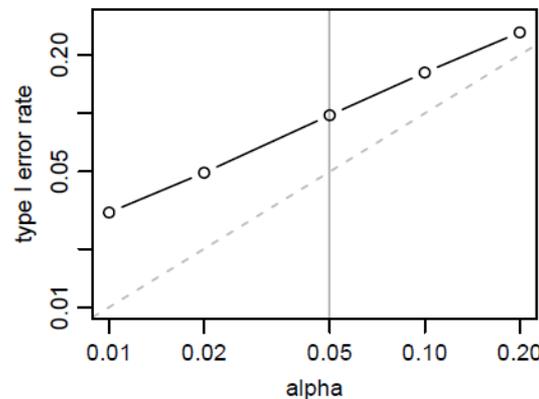
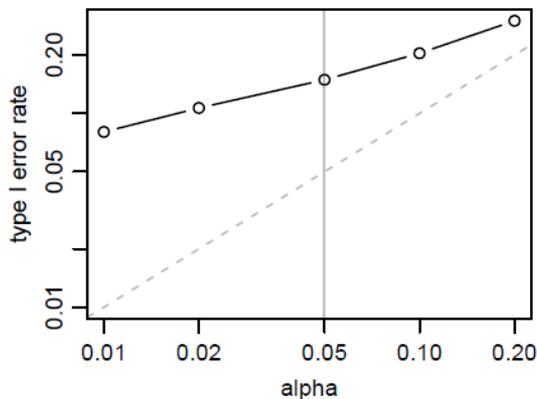
HKSJ



modified Knapp-Hartung



one large trial



MULTIPLICITY DUE TO MULTIPLE ENDPOINTS

- ▶ **Should we adjust for multiple primary endpoints?**
 - ▶ depends on question (see above)
 - ▶ Efficacy and safety endpoints on equal footing?
 - ▶ In RCTs: level not split between efficacy and safety
- ▶ **If yes, how?**
 - ▶ Technically problem largely unsolved ...

REQUIRED INFORMATION SIZE

- ▶ Application of group-sequential stopping boundaries (well-known from RCTs, e.g. O'Brien-Fleming alpha spending function) to cumulative meta-analyses
- ▶ “Required information size” (maximum information in group-sequential plan)
- ▶ Quote from (Step 4 in Jacobsen et al (2014)):
- ▶ “To estimate a required information size, it is necessary:
 - ▶ To estimate an anticipated intervention effect [...]
 - ▶ To estimate a variance of the anticipated difference in intervention effect [...]
 - ▶ To estimate a variance of the intervention effect estimates between trials [...]”

REQUIRED INFORMATION SIZE

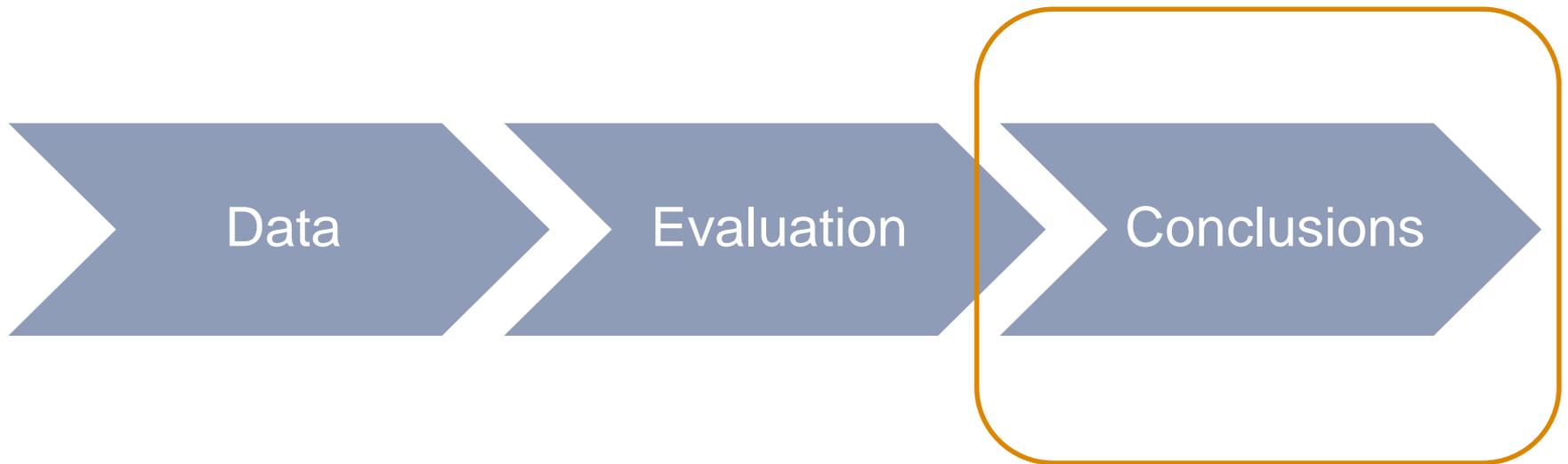
- ▶ Application of group-sequential stopping boundaries (well-known from RCTs, e.g. O'Brien-Fleming alpha spending function) to cumulative meta-analyses
- ▶ “Required information size” (maximum information in group-sequential plan)
- ▶ Quote from ...
- ▶ “To es ...
- ▶ To ...
- ▶ To es ...
intervention ...
- ▶ To estimate a variable ... the intervention effect estimates between trials [...]”

It would be tempting to do the systematic review / meta-analysis first ...

STATISTICAL SIGNIFICANCE AND CLINICAL RELEVANCE

- ▶ In the context with the interpretation of RCTs a number of suggestions made that could also be applied to meta-analyses
 - ▶ shifted null hypotheses to demonstrate clinically relevant differences (Victor, 1987)
 - ▶ Testing null hypothesis of no difference, but considering point estimate and confidence limits for relevance assessment (Jones, 2002)
 - ▶ Responder analyses (see e.g. Kieser et al, 2004)
 - ▶ Probabilistic index (relative effect) (see e.g. Kieser et al, 2013)

GENERATING EVIDENCE



CONCLUSIONS

- ▶ **Which data / studies to include / exclude?**
 - ▶ Moving away from dichotomy (in or out) to weighting evidence according to relevance
 - ▶ (Bayesian) generalized evidence synthesis (hierarchical models) as framework to be more inclusive
 - ▶ Accounting for between-study heterogeneity (and the uncertainty in estimating it)
- ▶ **Criteria for statistical significance / clinical relevance**
 - ▶ Statistical significance alone not sufficient, additional criteria reflecting clinical relevance necessary
- ▶ **Challenge: Balance both aspects sensibly**

SOME REFERENCES

- ▶ Friede T, Röver C, Wandel S, Neuenschwander B (2016) Meta-analysis of two studies in the presence of heterogeneity with applications in rare diseases. *Biometrical J* (in press)
- ▶ Friede T, , Röver C, Wandel S, Neuenschwander B (2016) Meta-analysis of few small studies in orphan diseases. *Research Synthesis Methods* (in press).
- ▶ Kieser M, Röhmel J, Friede T (2004) Planning considerations when assessing the clinical relevance of trial results by responder analyses. *Statistics in Medicine* 23: 3287-3305.
- ▶ Kieser M, Röhmel J, Friede T (2004) Power and sample size determination when assessing the clinical relevance of trial results by ‘responder analyses’. *Statistics in Medicine* 23:3287–3305.
- ▶ Kieser M, Friede T, Gondan M (2013) Assessment of statistical significance and clinical relevance. *Statistics in Medicine* 32: 1707–1719.
- ▶ Roberts I, Ker K, Edwards P, Beecher D, Manno D, Sydenham E (2015) The knowledge system underpinning healthcare is not fit for purpose and must change. *BMJ* 350, h2463.
- ▶ Röver C, Knapp G, Friede T (2015) Hartung-Knapp-Sidik-Jonkman approach and its modification for random-effects meta-analysis with few studies. *BMC Medical Research Methodology* 15: 99.
- ▶ Wegscheider K, Friede T (2016) Do we consent to rules of consent and confidentiality? (in preparation)