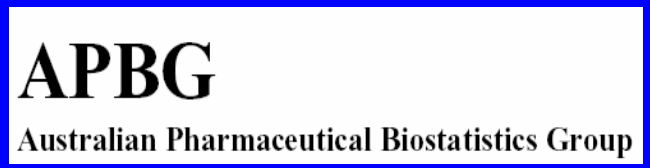
Indirect Comparisons Workshop

Introduction and Literature Review

Dr. Philip McCloud and Sandra Thompson
Roche Products Pty Ltd

30 March 2007



Motivation for Workshop

- Indirect comparisons contain subtle statistical issues
- Australian affiliates have little influence over comparators chosen for phase III trials by the parent company
- From the time of designing a trial, to applying for reimbursement, new comparators may enter the market thus forcing an indirect comparison at the PBAC decision point
- There may be many potential comparators, one cannot undertake direct comparisons against every single one
- Thus indirect comparisons are likely with all the consequences of bias, and poor power, which through no-ones fault may lead to ambiguous outcomes

Table of Contents

- Introduction
- Meta-analysis review
- Unadjusted or naive Indirect Comparisons
- Adjusted Indirect Comparisons

Adjusted Indirect Comparisons

- Fundamental Issues
  - Generalizability
    - Transitive fallacy, Consistency, Homogeneity of Treatment effect
  - Precision is out of control
  - Biased estimate of treatment effect
  - Fixed versus Random effects
    - Between or within study variance
  - Large sample requirements

Introduction

- One has always been told that cross study comparisons are flawed
- Why?
- IBT meeting – Pegasys versus PegIntron
- What has changed?

Pegasys versus PegIntron

	Fried (801) (Pegasys)	Hadz. (942) (Pegasys)	Mann (PegIntron)
All Genotypes	56%	63%	54%
Genotype-1	46%	52%	42%
Genotype non-1	76%	81%	82%

Unadjusted or Naive Indirect Comparison

SSAI
Statistical Society of Australia Inc.

APBG
Australian Pharmaceutical Biostatistics Group

Adjusted Indirect Comparison Bucher et al (1997)

- "Meta-analysis provides a potential solution to the problem of deciding between treatments that have not been directly compared."
- "Felson et al (1990) conducted a meta-analysis examining the relative effectiveness and toxicity of second-line drugs in rheumatoid arthritis."
- "...investigators chose to compare patients in different trials given competing second-line agents..."
- "Although all data in this meta-analysis came from randomised controlled trials, the power of randomisation is lost and the data are reduced to the equivalent of those derived from contemporaneous or historical cohort studies."

Adjusted Indirect Comparison Bucher et al (1997)

- So called unadjusted or naive indirect comparison; Bucher et al (1997) are setting the scene for the adjusted indirect comparison
- As Dr. Matei Popescu said at the IBT we cannot conduct cross study comparisons: read "unadjusted indirect comparisons", because the comparisons may be biased, and confounded with between study differences
- The estimate of the treatment effect remains biased unless the entire population of the meta-analysis is captured in one giant randomisation

WV16240: Study Design

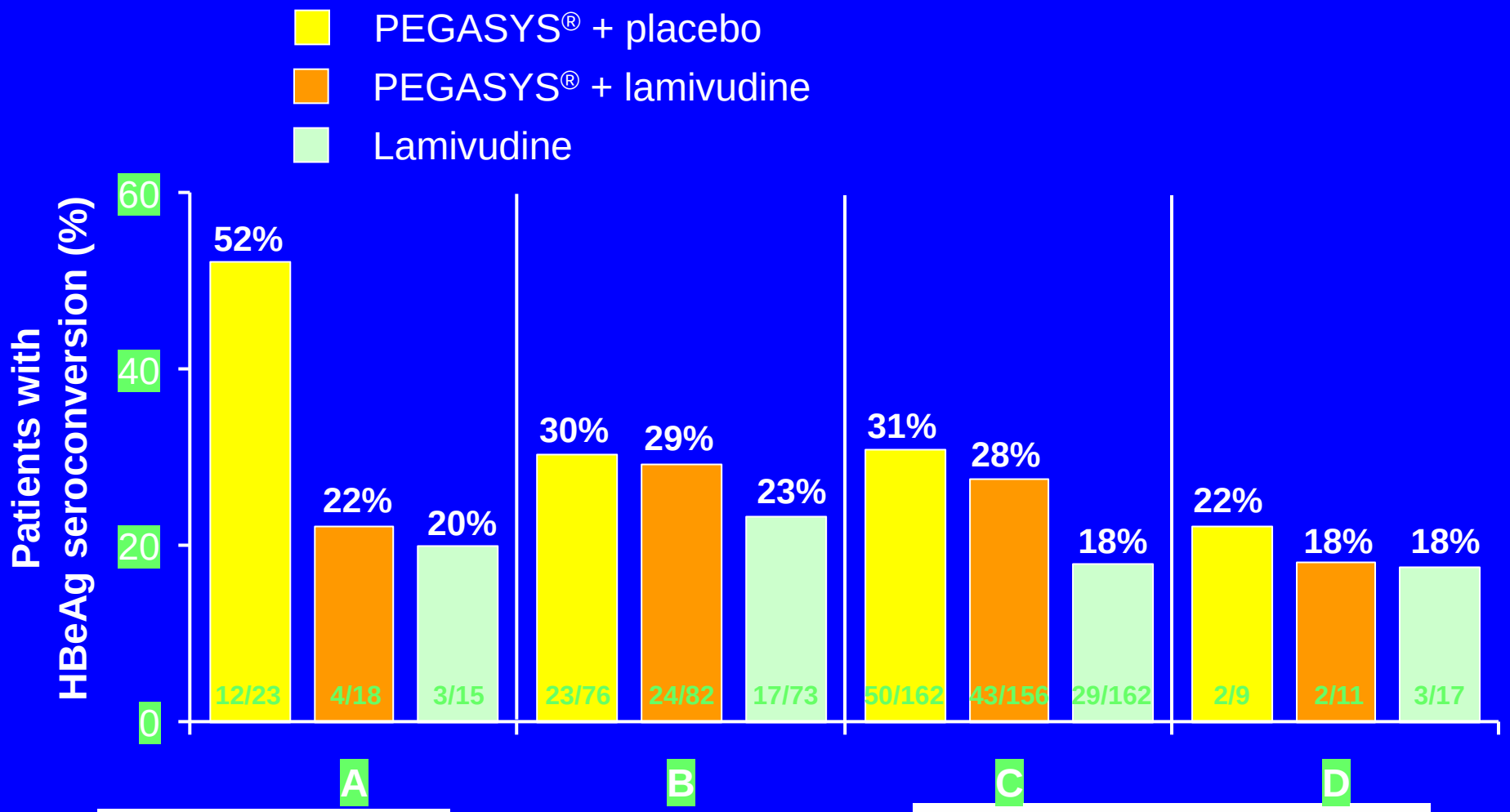
Patients with HBeAg-negative CHB were randomised using a 1:1:1 ratio

ITT population: n=814

SSAI Statistical Society of Australia Inc.

APBG Australian Pharmaceutical Biostatistics Group

HBeAg Seroconversion at EOF by Genotype



Statistical Society of Australia Inc.

APBG
Australian Pharmaceutical Biostatistics Group

Bias in Treatment Comparisons

- In conducting studies in Hepatitis B any cross-study comparisons are confounded by the proportion of patients from the respective genotypes
- The presence of different genotypes was not known until the mid-90's although people had observed the different seroconversion rates between Asia and Europe for many years
- Baseline characteristics of patient populations vary across studies thus compromising cross study comparisons

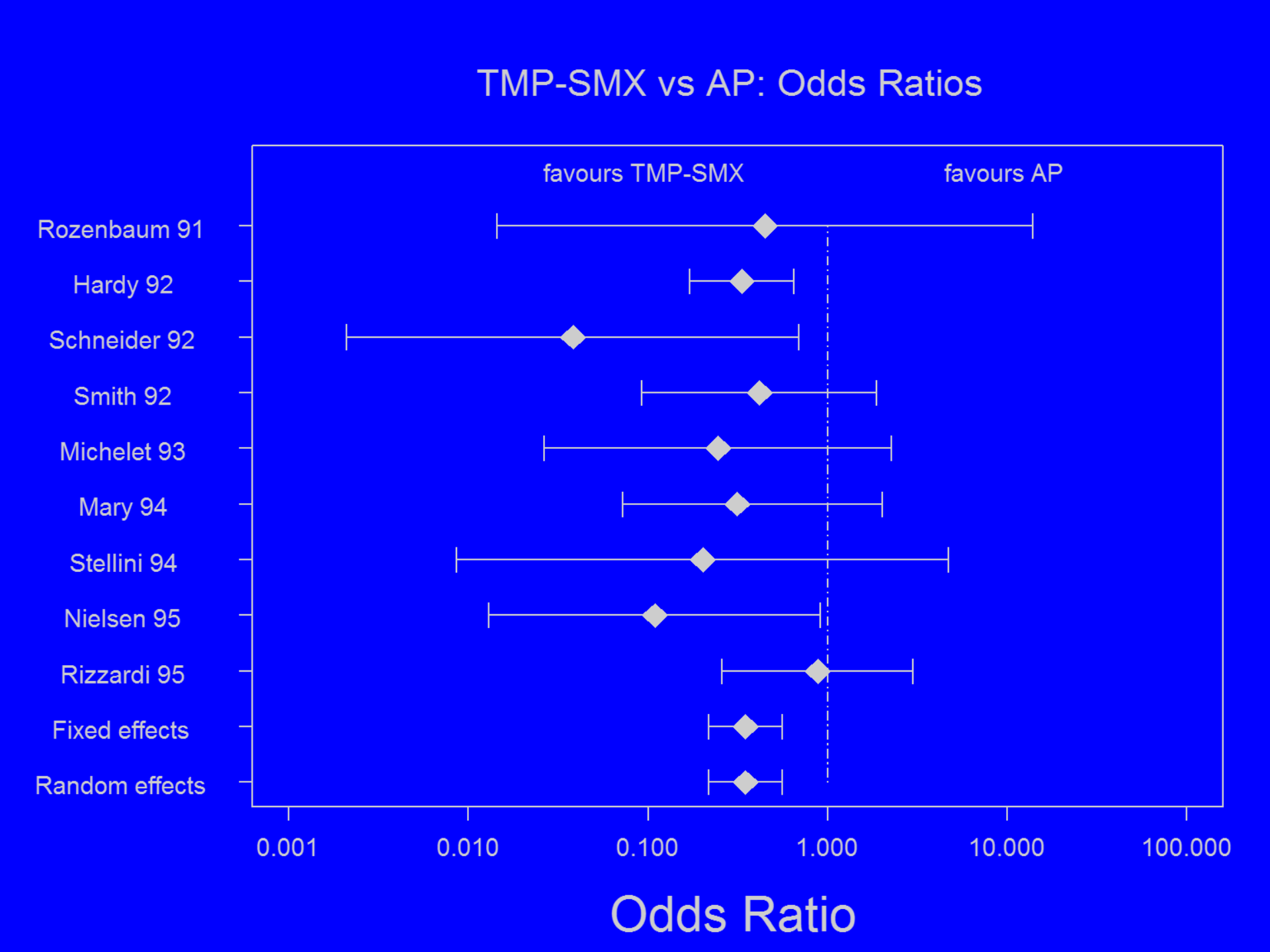
Meta-Analysis

SSAI
Statistical Society of Australia Inc.

APBG
Australian Pharmaceutical Biostatistics Group

Meta-Analysis

- Combine the results of several studies, which address the same research question
- Bucher et al (1997): two experimental, trimethoprim-sulphamethoxazole and dapsone/pyrimethamine, and one standard regimen, aerosolized pentamidine, for primary and secondary prevention of Pneumocystis carinii pneumonia in HIV infection
- Nine trials compared TMP-SMX with AP
- Primary efficacy: number of patients with Pcp


TMP-SMX vs AP: Odds Ratios
favours TMP-SMX
favours AP
Rozenbaum 91
Hardy 92
Schneider 92
Smith 92
Michelet 93
Mary 94
Stellini 94
Nielsen 95
Rizzardi 95
Fixed effects
Random effects
0.001
0.010
0.100
1.000
10.000
100.000
Odds Ratio

Statistical Details

- Let Y_i denote the estimator of the effect in study *i*
- Let θ_i denote the parameter of interest for study *i;* eg odds-ratio, relative risk, hazard ratio
- The fixed effect model assumes that $\theta_i = \theta$ for all *i*; that is the true value of the parameter of interest is constant across studies

Statistical Details

- The meta-analysis estimate for θ is obtained with weights proportion to the reciprocal of the variance in each study

$$\hat{\theta} = \frac{\sum_i W_i Y_i}{\sum_i W_i}$$

with se

$$\left(\sum_i W_i \right)^{-\frac{1}{2}}$$

- Note that the estimates of θ and its se do not require any large sample assumptions

Justification for the fixed effect analysis

- Senn gave three:
- (1) The null hypothesis tests $\theta_i = 0$ for all *i* against the alternative hypothesis that

  $\theta_i \neq 0$ for at least one *i*
- We often make such an assumption in conducting hypothesis tests; ie assume equality to zero, and therefore assume homogeneity; we now look for strong evidence of a contradiction

Justification for the fixed effect analysis

- (2) The treatment effect, θ , is estimated within each study. Therefore the combined estimate of, θ , can be tested against a combined estimate of within study variance
- (3) We believe that the treatment effect is homogeneous, that is $\theta_i = \theta$ for all *i*

Testing the homogeneity

- **Cochran's Q statistic:**

$$Q = \sum_i W_i \left(Y_i - \hat{\theta} \right)^2$$

which follows a chi-square distribution on *k-1* degrees of freedom

Testing the homogeneity

- Senn comments:
- It is sometimes claimed that the reason for doing a random-effects meta-analysis is that this assumption has been proved false
- However this is rather misleading because:
  - The other two justifications for a fixed effects analysis still remain
  - In practice, if the fixed effects meta-analysis rejects the null hypothesis that the treatment effects are zero, then nobody believes the assumption (which assumption?)

Testing the Homogeneity

- In the presence of significant heterogeneity how you proceed depends on whether you are a conditional person or a marginal person
- A conditional person will examine the differences between within study treatment estimates in search of an explanation
- A marginal person will argue that they are only interested in the combined treatment effect
- Public health has been advanced with both approaches

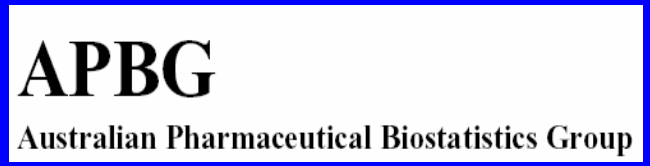
Summary

- The same Treatment effect is estimated in a series of studies
- For each study the Treatment effect can be tested against within study variance
- Therefore the combined Treatment effect can be tested against a combined estimate of within study variance
- No large sample assumption is required to justify the test

Summary

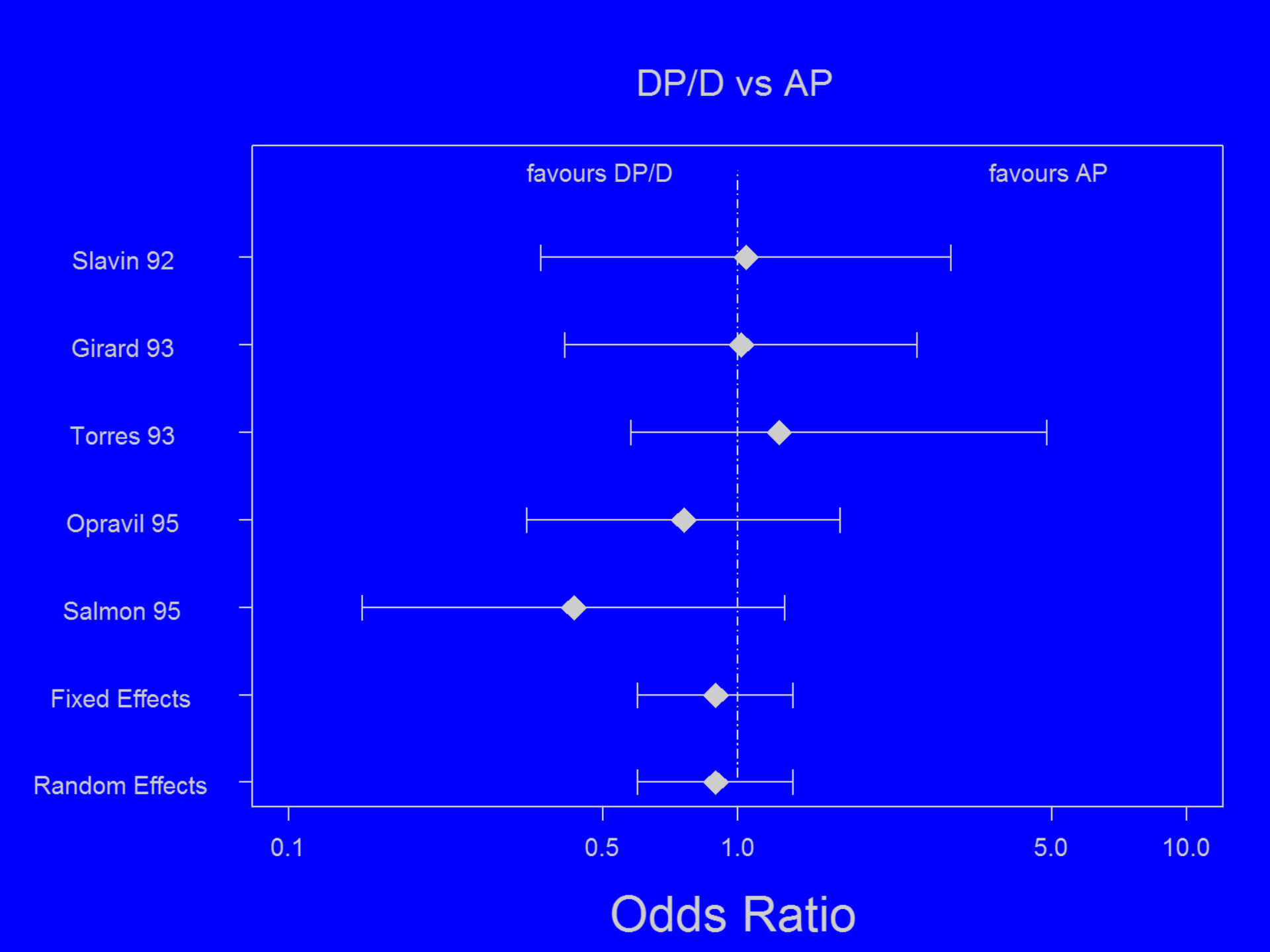
- Within each study the randomisation provides an unbiased estimate of the treatment effect and variance
- In my opinion we are too quick to discard the fixed effect analysis in favour of the random effects in the classical meta-analysis setting: "similar protocol, same treatments, similar populations"

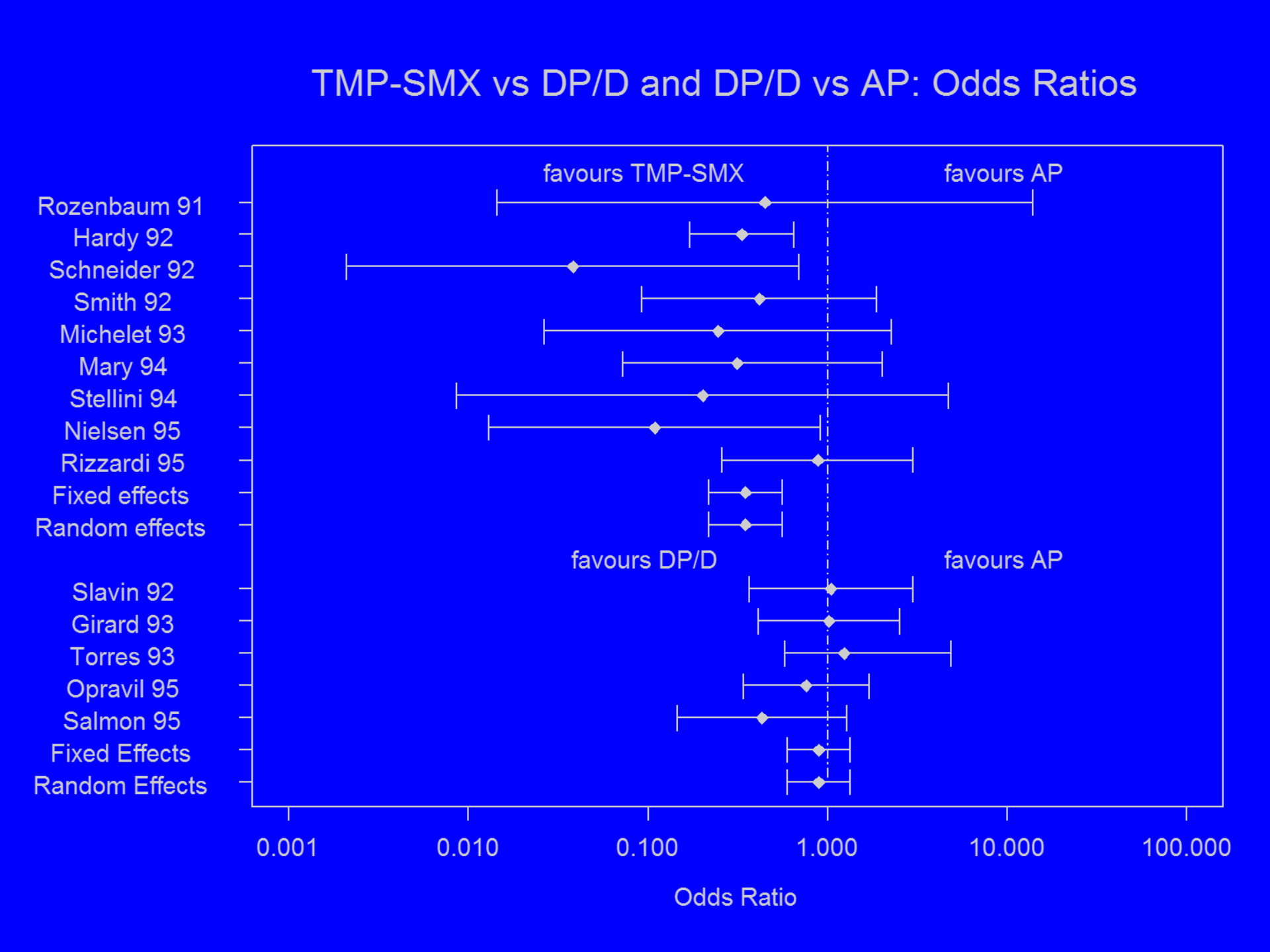
Adjusted Indirect Comparison



Adjusted Indirect Comparison Bucher et al (1997)

- "...is there any alternative strategy for pooling data that is less susceptible to bias?"
- "...examine the magnitude of the treatment effect (ie treatment vs placebo) in studies of different treatments and compare those treatment effects. ...we MAY obtain an unbiased estimate of the treatment effect."
- Five trials compared Daspsone/pyrimethamine with aerosolised pentamidine
- Primary efficacy parameter was Pneumocystis carinii pneumonia (Pcp)


DP/D vs AP
favours DP/D
favours AP
Slavin 92
Girard 93
Torres 93
Opravil 95
Salmon 95
Fixed Effects
Random Effects
0.1
0.5
1.0
5.0
10.0
Odds Ratio


TMP-SMX vs DP/D and DP/D vs AP: Odds Ratios
favours TMP-SMX
favours AP
Rozenbaum 91
Hardy 92
Schneider 92
Smith 92
Michelet 93
Mary 94
Stellini 94
Nielsen 95
Rizzardi 95
Fixed effects
Random effects
favours DP/D
favours AP
Slavin 92
Girard 93
Torres 93
Opravil 95
Salmon 95
Fixed Effects
Random Effects
0.001
0.010
0.100
1.000
10.000
100.000
Odds Ratio

Adjusted Indirect Comparison Bucher et al (1997)

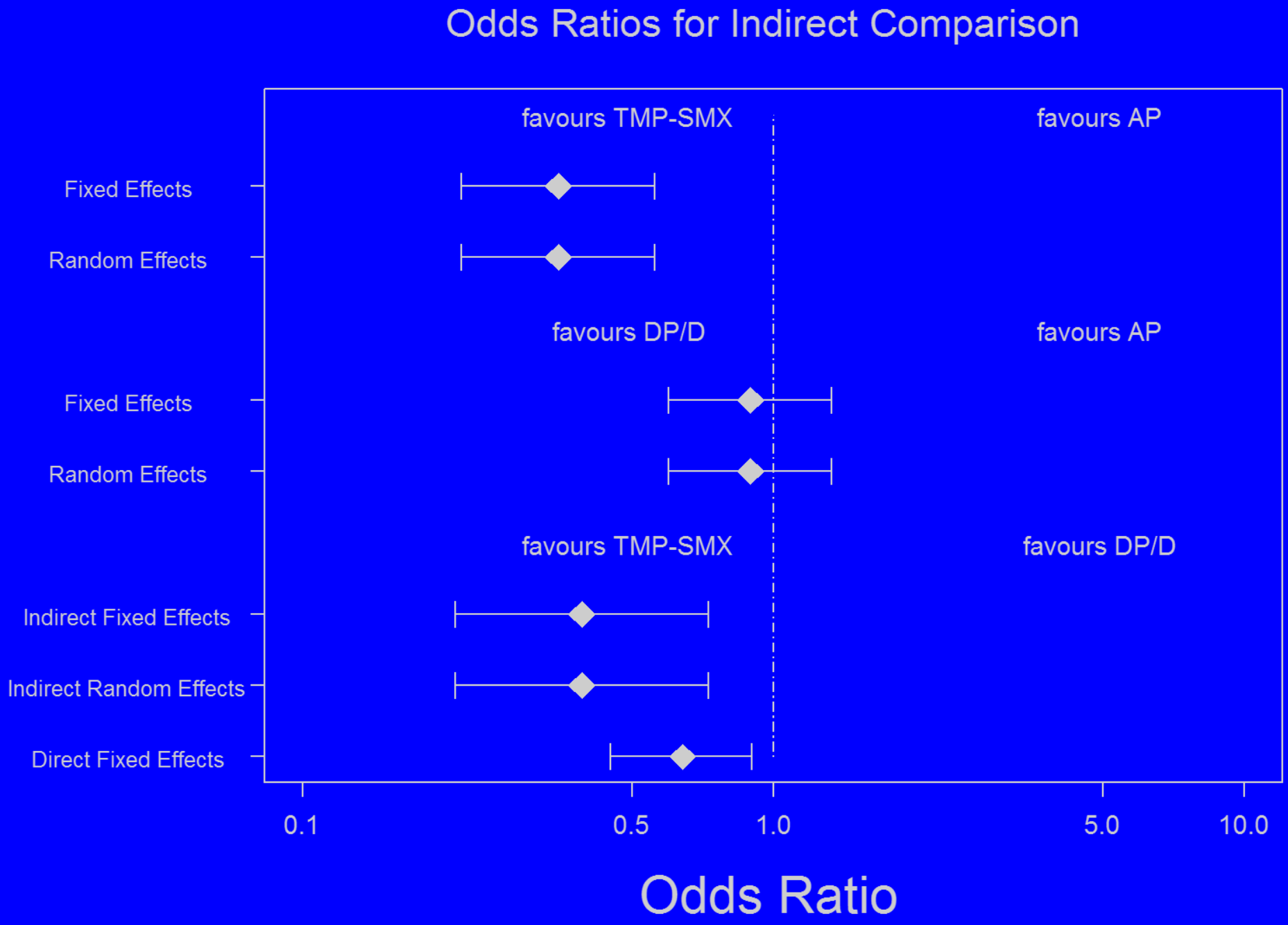
- Treatment effect of TMP-SMX versus AP with standard error, se(TMP-SMX vs AP), and Treatment effect of DP/P versus AP with standard error, se(DP/P vs AP).
- Adjusted Indirect Comparison of TMP-SMX versus DP/P is achieved by subtracting the two treatment effects, which will have a variance of the sum of var(TMP-SMX vs AP) and var(DP/P vs AP)

Adjusted Indirect Comparison Bucher et al (1997)

- TMX-SMP vs AP
  - OR = 0.349; LCL = 0.217; UCL = 0.560
  - Cochran Q = 5.92 on 8 df
  - Log(OR) = -1.053; se = 0.241
- DP/P vs AP
  - OR = 0.892; LCL = 0.598; UCL = 1.329
  - Cochran Q = 2.76 on 4 df
  - Log(OR) = -0.115; se = 0.204

Adjusted Indirect Comparison Bucher et al (1997)

- Adjusted indirect comparison
  - $Log(OR_{TMP/AP}) - log(OR_{DP/AP}) = -1.053 + 0.115$
  - Log(OR) = -0.938
  - Var(diff) = se^2 + se^2 = 0.0997
  - Se(diff) = 0.316
  - Cochran Q = 5.92 + 2.76 = 8.68 on 12 df
  - OR_{diff} = 0.39; LCL = 0.211; UCL = 0.726
- Direct comparison of TMP-SMX with DP/P
  - OR = 0.64; LCL = 0.45; UCL = 0.90


Odds Ratios for Indirect Comparison
favours TMP-SMX
favours AP
Fixed Effects
Random Effects
favours DP/D
favours AP
Fixed Effects
Random Effects
favours TMP-SMX
favours DP/D
Indirect Fixed Effects
Indirect Random Effects
Direct Fixed Effects
0.1
0.5
1.0
5.0
10.0
Odds Ratio

Generalisability

Adjusted indirect comparison illustrated by antimicrobial prophylaxis Song et al (2000)

- Generalizability: "For the adjusted indirect comparison to be valid, the key underlying assumption is that the relative efficacy of an intervention is consistent in patients included in different trials. That is the estimated relative efficacy should be generalizable, an issue of external validity."

Adjusted indirect comparison illustrated by antimicrobial prophylaxis Song et al (2000)

- "Generalizability of trials is often questionable because of restricted inclusion criteria, exclusion of patients, and higher level of settings..."
- "Because of many factors that may influence the generalizability of trials, it will often be difficult to decide whether the basic assumption of an adjusted indirect comparison is met."
- Elsewhere Song et al (2000) refer to this as a STRONG assumption

Indirect comparisons of competing interventions, Glenny et al (2005)

- "The key additional assumption of an indirect comparison using the results of trials of A vs B and A vs C is that there should be no important differences between the two sets of trials with respect to aspects that could influence (bias) the estimated treatment effect of B vs C, that is, there is no confounding of the comparison by some trial characteristic."
- Available from www.ncchta.org

Fisher et al (Am. Heart J, 2003)
How would a new agent compare with placebo?

- Clopidogrel and aspirin were compared in the CAPRIE study of patients with a recent MI, ischemic stroke, or symptomatic peripheral arterial disease. 40 trials comparing aspirin to placebo were summarised by the Antiplatelet Trialists' Collaboration.
- Compare Clopidogrel to placebo with an adjusted indirect comparison

Fisher et al (Am. Heart J, 2003)
How would a new agent compare with placebo?

- "It is clear that a major assumption is made here. It is assumed that the data for the active-control versus placebo estimate the same OR as that that would have occurred in the active-control study, had a concurrent placebo arm been included. Because the active-control will usually have been established earlier, the data for the active-control versus placebo is historic. Thus, a historic control is used."

Fisher et al (Am. Heart J, 2003)
How would a new agent compare with placebo?

- At a USFDA meeting "several USFDA statisticians expressed concern about the assumptions inherent in using the historic randomised clinical trials for the active control. They voiced concern that the treatment effect in current trials might be much less than in the prior trials..."
- Note concern regarding the changing treatment effect over time thus rendering the adjusted indirect comparison unreliable.
- Patient populations or outcomes change as new agents enter the market (Bonviva)

Fisher et al (Am. Heart J, 2003)
How would a new agent compare with placebo?

- "The placebo-controlled data are historic...with a constantly changing background...Therefore in addition to any statistical uncertainty...there is the uncertainty associated with the use of historic active-control / placebo data and different populations. This uncertainty is not generally amenable to estimation, so scientific judgment, expertise, and experience must play a role." cf Bucher et al (1997)

Smith & Kramer (BioMed, 2002) The transitive fallacy

- Use a graphical method to demonstrate the transitive fallacy of clinical trials
- Let A, B, and C be three statements. In formal logic, if A implies B, and B implies C, then A implies C. This is an example of transitivity.
- This logic does not extend to the design and interpretation of clinical trials, because statistical association is not generally transitive. (Simpson's paradox)

Smith & Kramer (BioMed, 2002) The transitive fallacy

- Let A denote radical gastrectomy & splenectomy, B denote "simple" gastrectomy, and C denote radiation
- The endpoint is survival, and the unobservable covariate is supportive care, which will improve overtime with better intensive care, better antibiotics, etc
- Hence an early trial of B versus C will favour C
- A later trial of A versus B shows similar results

Smith & Kramer (BioMed, 2002) The transitive fallacy

- If transitivity held then C should better A, however, the improved supportive care may even reverse the result
- Conclusion: "Validity of the sequential studies strategy (B versus C, and A versus B) rests on the assumption that there is no intervening important covariate that could confound the implied principle of transitivity."

Precision is out of control

SSAI
Statistical Society of Australia Inc.

APBG
Australian Pharmaceutical Biostatistics Group

Adjusted Indirect Comparison Bucher et al (1997)

- The efficacy of the adjusted indirect comparison is poor compared to a direct comparison
- "...in practice meta-analyses are carried out on existing studies, so the precision of the estimated treatment effect is not under the control of the analyst..."

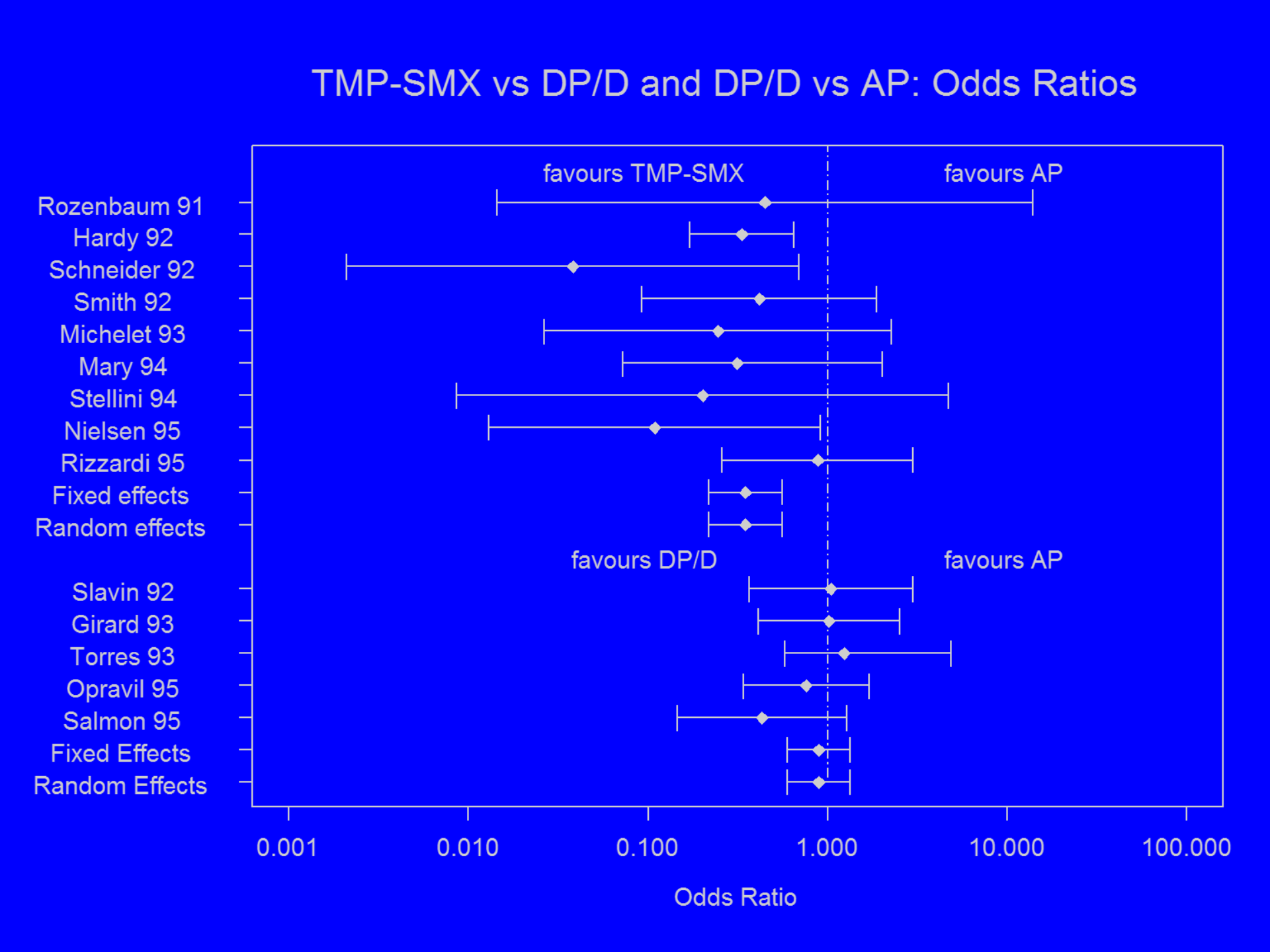
Adjusted Indirect Comparison Bucher et al (1997)

- "Thus our findings suggest that the strength of inference associated with indirect comparisons is inevitably limited, even in the absence of demonstrable sources of bias."
- "Even in the absence of evident differences, the strength of inferences from indirect comparisons is limited..."

Fixed versus Random effects
Between or within study variance

SSAI
Statistical Society of Australia Inc.

APBG
Australian Pharmaceutical Biostatistics Group


TMP-SMX vs DP/D and DP/D vs AP: Odds Ratios
favours TMP-SMX
favours AP
Rozenbaum 91
Hardy 92
Schneider 92
Smith 92
Michelet 93
Mary 94
Stellini 94
Nielsen 95
Rizzardi 95
Fixed effects
Random effects
favours DP/D
favours AP
Slavin 92
Girard 93
Torres 93
Opravil 95
Salmon 95
Fixed Effects
Random Effects
0.001
0.010
0.100
1.000
10.000
100.000
Odds Ratio

Adjusted Indirect Comparison Bucher et al (1997)

- What is different about the forest plot of TMX-SMP vs AP and DP/P vs AP?

Adjusted Indirect Comparison Bucher et al (1997)

- The key difference is that treatments TMP-SMX and DP/P are not compared within the same study:
  - Thus within study variance is not appropriate for the test of the treatment effect of TMP-SMX versus DP/P
  - The estimate of the treatment effect of TMP-SMX versus DP/P, and the estimate of within study variance cannot be assumed unbiased since randomisation is not at the appropriate level

Adjusted Indirect Comparison Bucher et al (1997)

- "The only requirement is that the magnitude of the treatment effect is constant across the differences in the populations' baseline characteristics."
- How do we know this is true? Should always use between study variance to test the adjusted indirect comparison, and
- Further, cannot assess the homogeneity of the two treatments that are being compared indirectly

Indirect comparisons of competing interventions, Glenny et al (2005)

- "Fixed effect direct and indirect methods do not give the right coverage: the proportion of occasions when the 95% confidence intervals do not include the correct value (i.e. the direct estimate based on all the trials) exceeds 5% as a consequence of the excess heterogeneity. By contrast, for random effects models the coverage is about 5%. ... The coverage of the naive method is awful, with over 40% of confidence intervals not including the correct value."

Indirect comparisons of competing interventions, Glenny et al (2005)

- "...an indirect comparison does not require homogeneity, and modelling approaches exist that include random effects terms that estimate the degree of heterogeneity in the comparisons."

Large Samples

Adjusted Indirect Comparison Bucher et al (1997)

- "The estimate...from this approach will be unbiased in LARGE SAMPLES if there is no interaction between covariates defining subgroups of patients...and the magnitude of the treatment."
- Note that "large samples" means a large number of trials, not simply a large number of patients (n > 20)

Lim et al (BMJ, 2003)
Compared Medium to Low dose aspirin

- Adjusted indirect comparison of medium dose versus low dose aspirin for prevention of graft occlusion
- Only five trials were included in the meta-analysis, and the adjusted relative risk ratio of 0.74, 95% CI = (0.52, 1.06) was inconclusive
- Only fixed effects was used since the trial results were homogeneous
- Concern about estimating variance components on less than five df

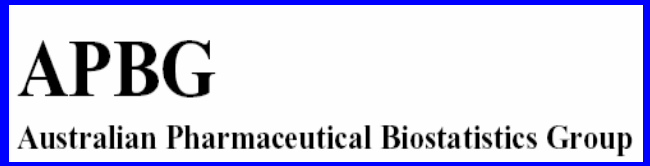
Indirect comparisons of competing interventions, Glenny et al (2005)

- "...in most situations there are too few trials for each paired comparison to allow reliable assessment of whether there is excess heterogeneity, or estimation of separate random effects for each component comparison."
- In chapter 5, Glenny et al (2005): "Analyses were performed for k=8,4,2, or 1, but some methods could not be used reliably for small k."
- In my opinion if we have too few trials then we should not undertake a formal indirect comparison

What is large?

- A historic rule of thumb has been that for greater than 30 independent observations large sample theory applies
- Prof Roger Mead commented: minimum of 12 better if greater than 20
- Uncomfortable with < 5

Empirical Comparisons



Adjusted indirect comparison illustrated by antimicrobial prophylaxis, Song et al (2000)

- "...identified 11 sets of trials in which different antibiotics can be compared both directly and indirectly." Method of Bucher used for the adjusted indirect comparison.
- "...considerable discrepancies exist between the direct and indirect comparisons..."
- "...even the adjusted indirect comparison may provide an estimate that is very different from the direct comparison within randomised controlled trials..."

Adjusted indirect comparison illustrated by antimicrobial prophylaxis, Song et al (2000)

- "...the adjusted method of indirect comparison may not give results similar to the direct comparison."
- "...the results of any indirect comparisons should be interpreted with great caution."

Empirical assessment
Song et al (BMJ, 2003)

- Song et al (2003) conducted an empirical assessment of indirect comparisons
- Method: searched the database of abstracts of reviews of effectiveness, Cochrane database, Medline, and reference
- Results: identified 28 systematic reviews in which both direct and in direct comparisons could be conducted; in total 44 comparisons

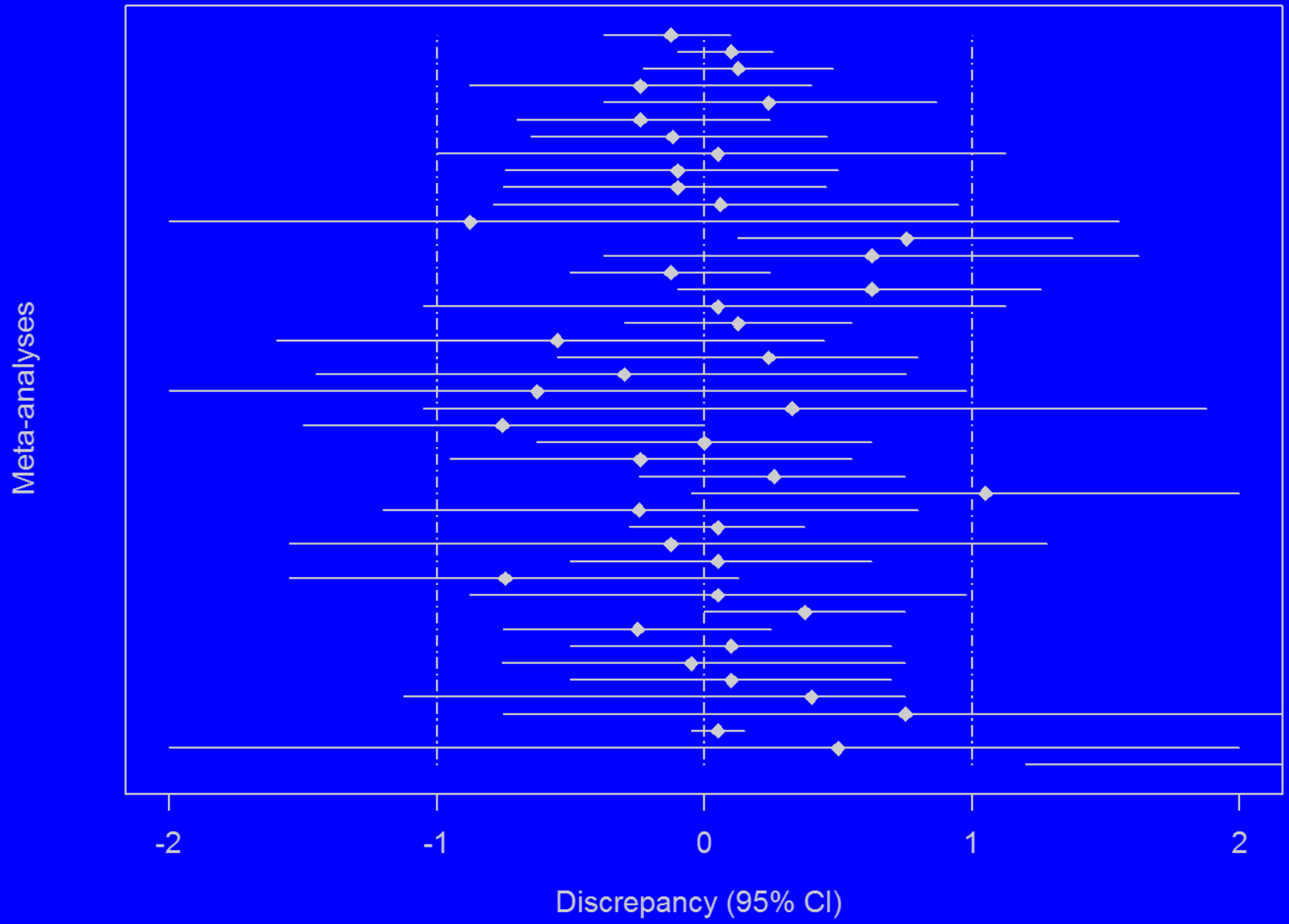
Empirical assessment
Song et al (BMJ, 2003)

- Two measures of discrepancy between direct estimate (T_D) and adjusted indirect estimate (T_{AI}):
  - $\Delta = T_D - T_{AI}$ with $SE(\Delta) = SQRT(SE(T_D)^2 + SE(T_{AI})^2)$
  - $Z = \Delta / SE(\Delta)$

Empirical assessment
Song et al (BMJ, 2003)

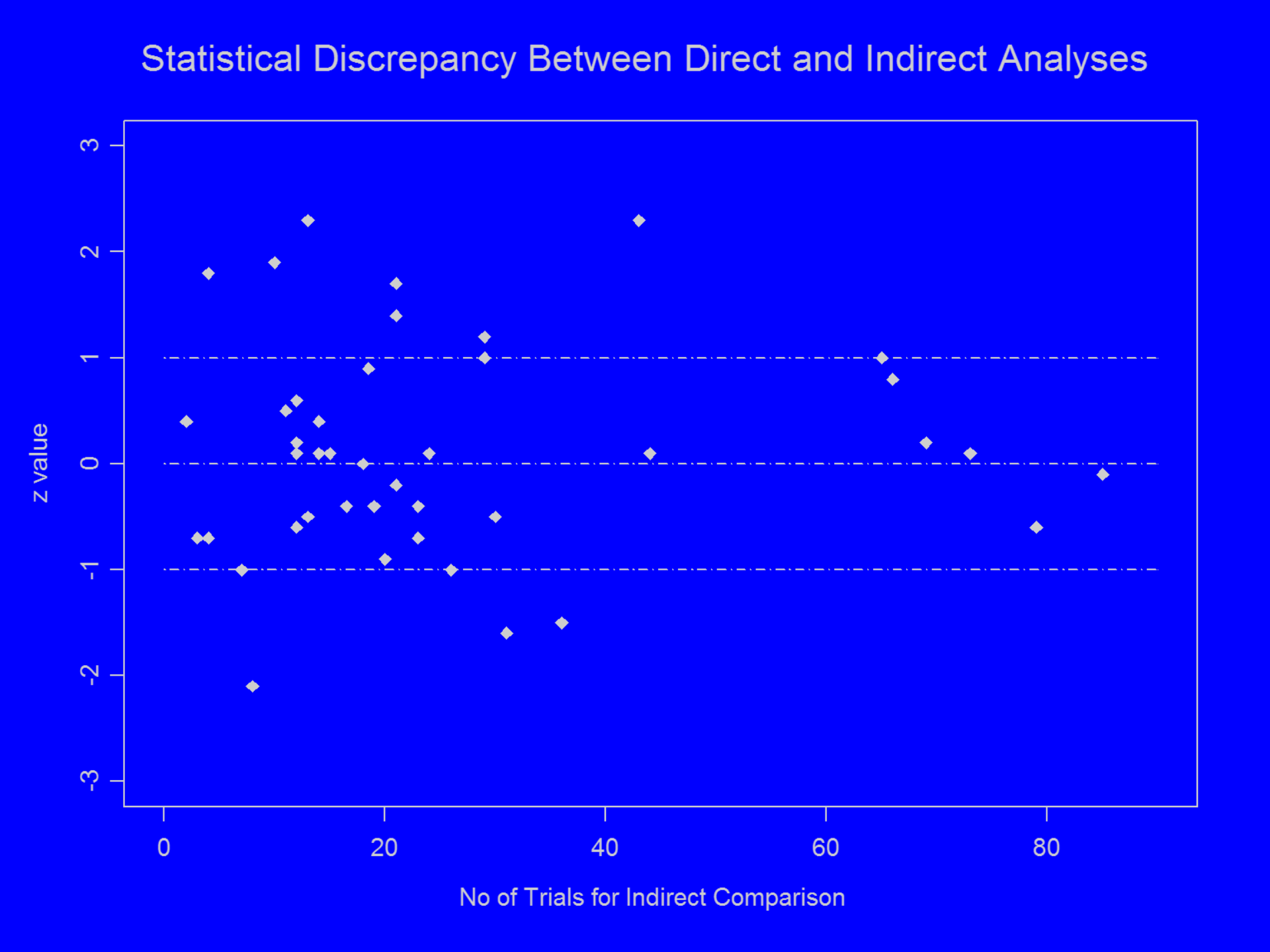
- Figure below summarises the discrepancies between direct and adjusted indirect comparisons
- Three of the 44 comparisons were significant at 0.05 and 4 at 0.1; the 41 NS results prove little given the likely low power
- Relative efficacy was equally likely to be over or under estimated by the indirect compared to the direct comparison

Discrepancy Between Direct and Adjusted Indirect Comparison



Empirical assessment
Song et al (BMJ, 2003)

- Figure below shows that "statistical discrepancies, as measured by z, tended to be smaller when the number of trials was large (>60) than when the number of trials was small (<40)"
- Note the very large number of trials for the adjusted indirect comparisons; cf large sample result mentioned by Bucher et al (1997)
- Note 15 comparisons with Z ≥~ 1.0 SE; that is one-third of the comparisons ≥~ 1.0 SE; very unreliable


Statistical Discrepancy Between Direct and Indirect Analyses
z value
3
2
1
0
-1
-2
-3
0
20
40
60
80
No of Trials for Indirect Comparison

Empirical assessment
Song et al (BMJ, 2003)

- "Compared with the direct estimate the adjusted indirect comparison estimates were less likely to be significant. Ten of the 19 significant direct estimates became non-significant in the adjusted indirect comparison, while only two of the 25 non-significant direct estimates was significant in the adjusted indirect comparison."
- On the empirical evidence we loose slightly more than 50% of the significant results; too unreliable even with a large number of trials

Empirical assessment
Song et al (BMJ, 2003)

- "Partly because of the wide confidence interval provided by the adjusted indirect comparison, significant discrepancies between the direct and the adjusted indirect estimate were infrequent (3/44)."

Conclusion

- Concluded: “Our example demonstrates that while this new model protects against some biases, it may still lead to an inaccurate estimate of treatment effect.” Bucher et al (1997)
- Many warnings and caveats about the dangers of adjusted indirect comparisons which are easily forgotten

Thank you for your attention !!

SSAI
Statistical Society of Australia Inc.

APBG
Australian Pharmaceutical Biostatistics Group