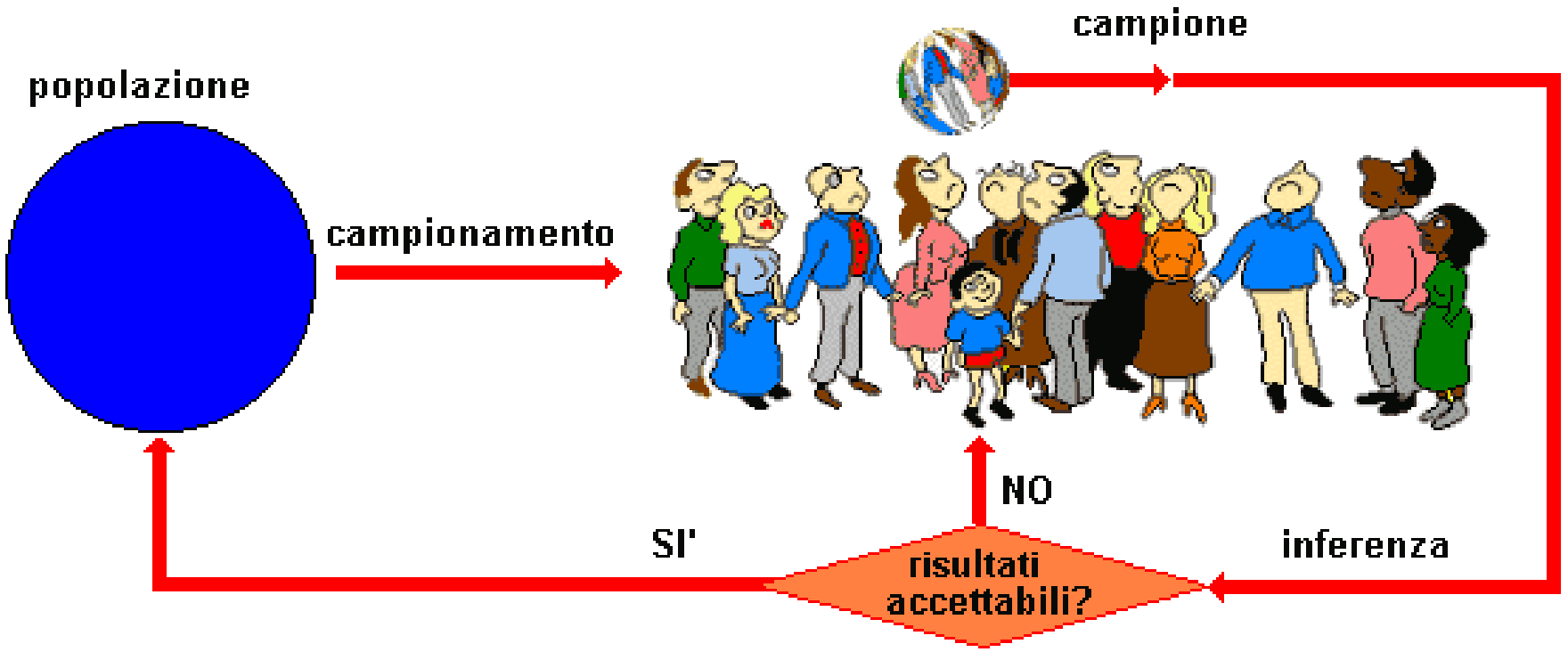
Imprecision,
Multiplicity,
analisi per sottogruppo


popolazione
campionamento
campione
NO
SI'
risultati accettabili?
inferenza

Uncertainty Estimation

- When we measure some physical quantity with an instrument and obtain a numerical value, we want to know how close this value is to the true value. The difference between the true value and the measured value is the error. Unfortunately, the true value is unknown and unknowable. If we knew it, we wouldn't need the experiment. Since this is the case, the exact error is never known. We can only estimate it.

Imprecision

- Gli errori casuali condizionano la ***precisione della stima campionaria***

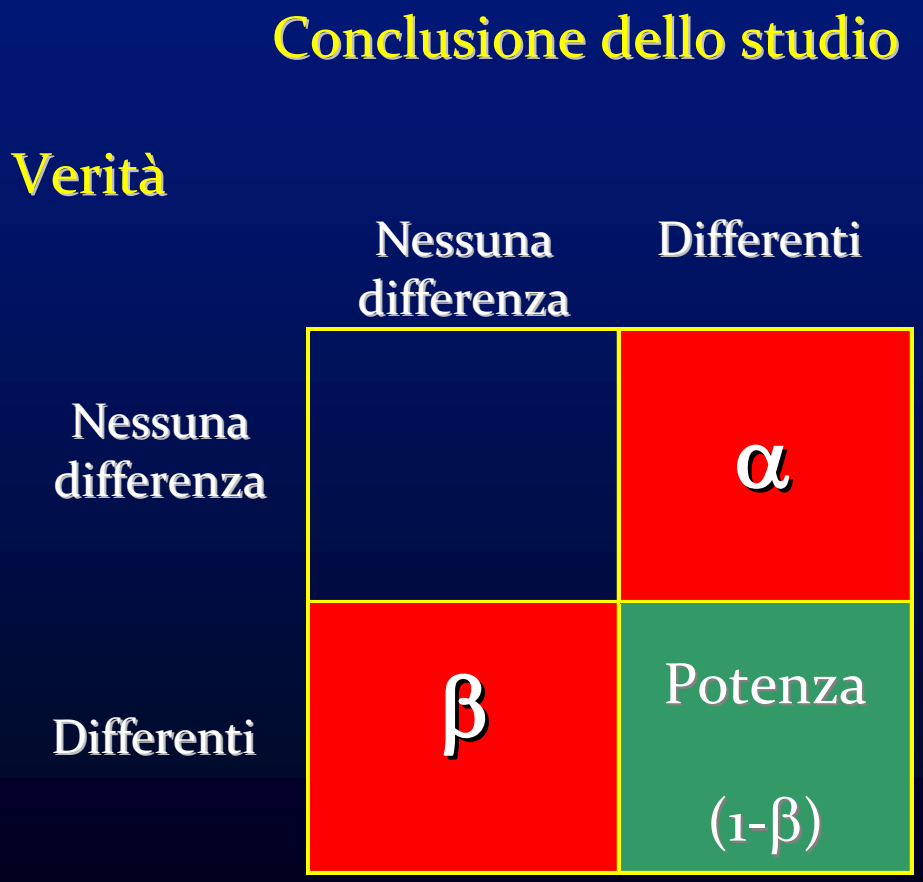
Imprecision

- ***Il controllo della variabilità casuale deve essere effettuato:***
  - ***In fase di pianificazione dello studio – minima dimensione campionaria sufficiente per saggiare l'ipotesi nulla***
  - ***In fase di analisi - accompagnando la stima puntuale da una misura della sua variabilità casuale***

Tipi di errore

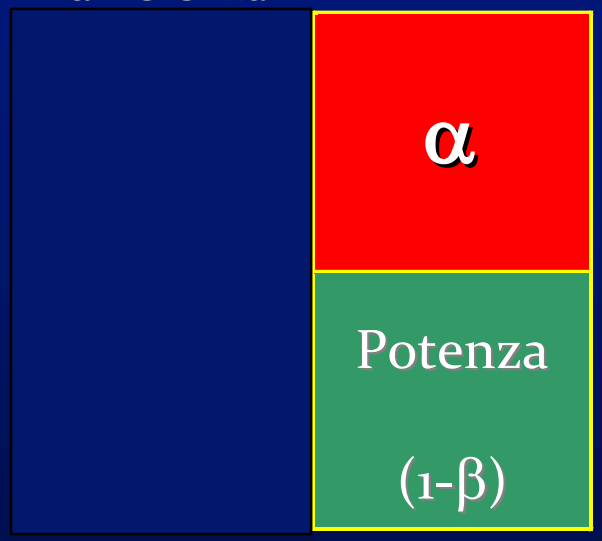
In fase di pianificazione devo tenere conto dei 2 possibili errori che potrò compiere

In fase di interpretazione dei risultati devo confrontarmi con solo uno dei due



Conclusione dello studio

Verità

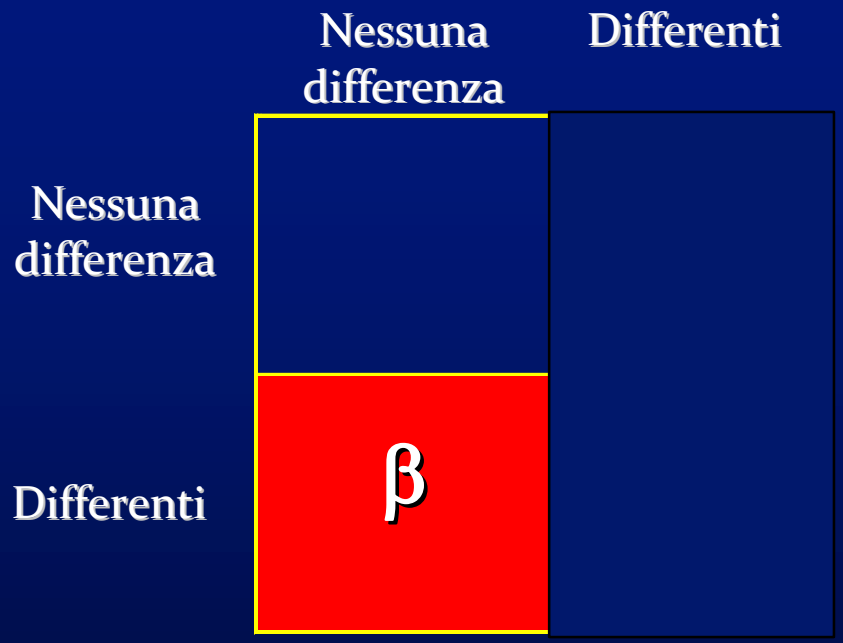


Interpretazione

Se la mia conclusione è "ESISTE DIFFERENZA" posso commettere solo un errore di I tipo, che è quantificato dal valore di p

Conclusione dello studio

Verità

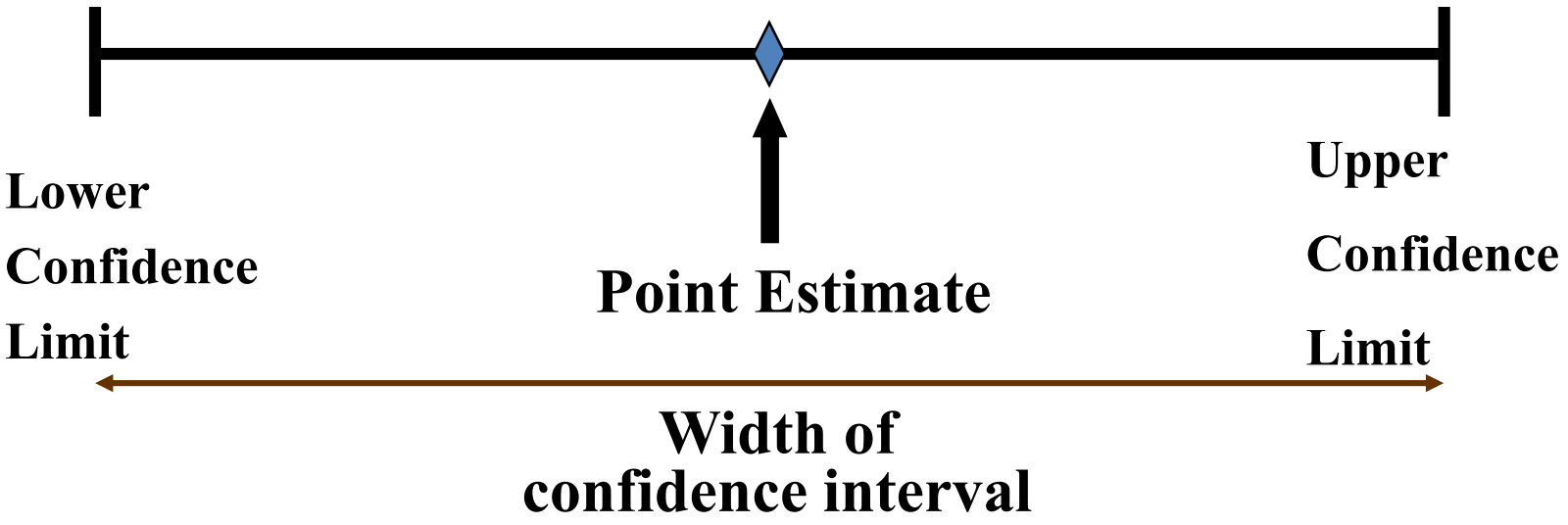


Interpretazione

Se la mia conclusione è "NON ESISTE DIFFERENZA" posso commettere solo un errore di II tipo, che quantifico calcolando la potenza del test

imprecision

- INTERVALLO DI CONFIDENZA



Imprecision

- small sample size
  - small number of events

Imprecision

Dependent on the choice of the difference (Δ) you wish to detect and the resulting sample size required

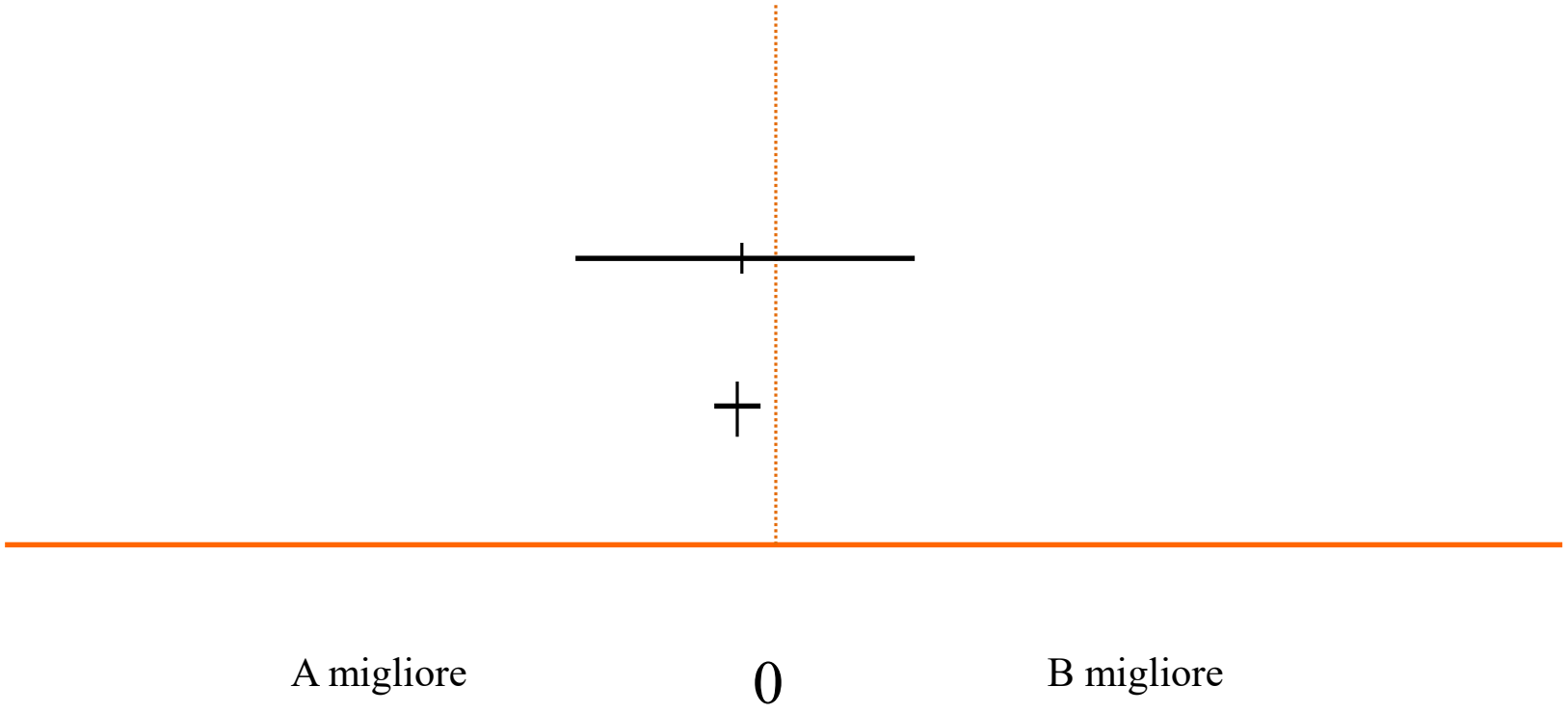
Significatività statistica e rilevanza clinica

- Se con un'opportuna dimensione del campione siamo in grado di ottenere risultati significativi, ciò non ci permette ancora di capire quanto essi lo siano dal punto di vista clinico

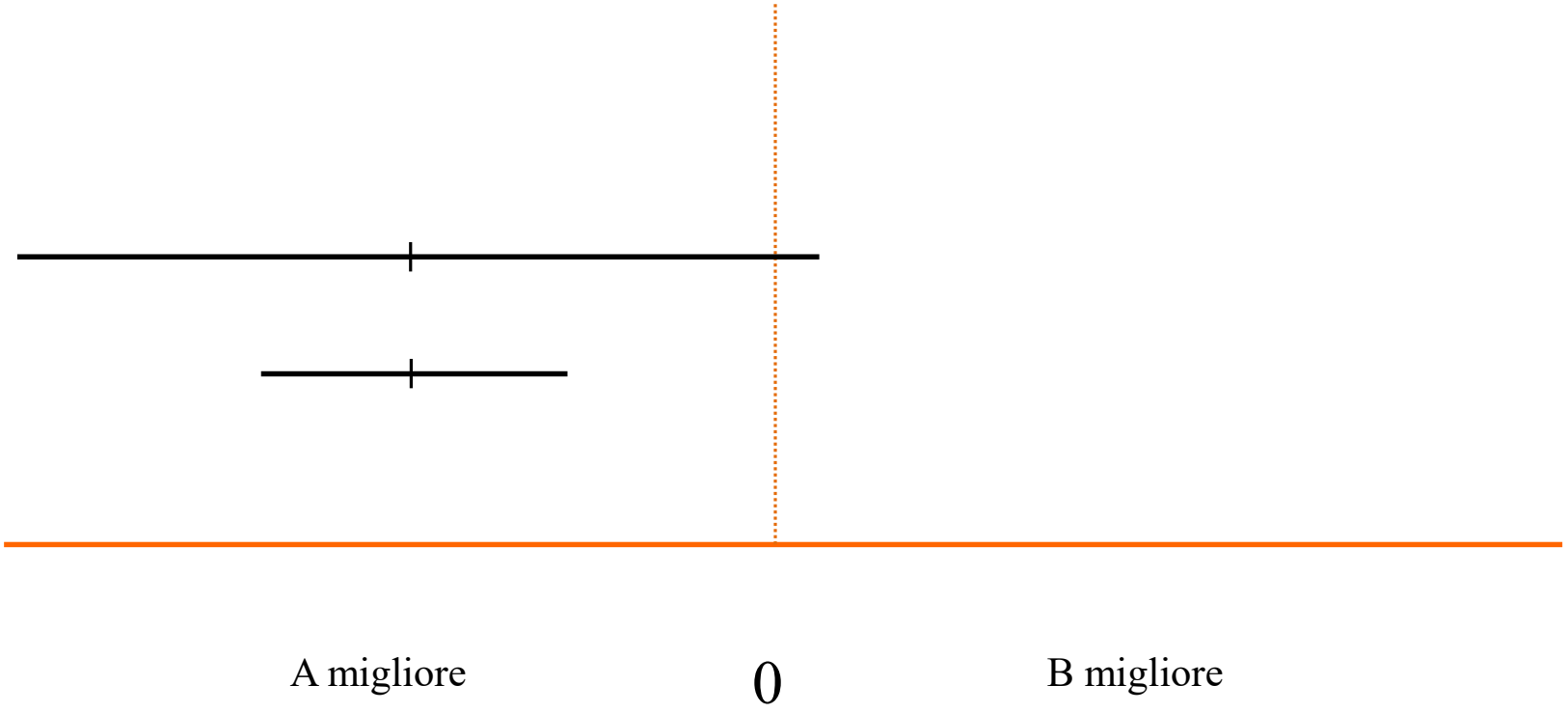
- $P<0.05$ potrebbe includere differenze clinicamente irrilevanti

- $P\geq0.05$ potrebbe nascondere una differenza reale ed importante, che non è stata evidenziata a causa di una bassa potenza

# P<0.05 potrebbe includere differenze clinicamente irrilevanti

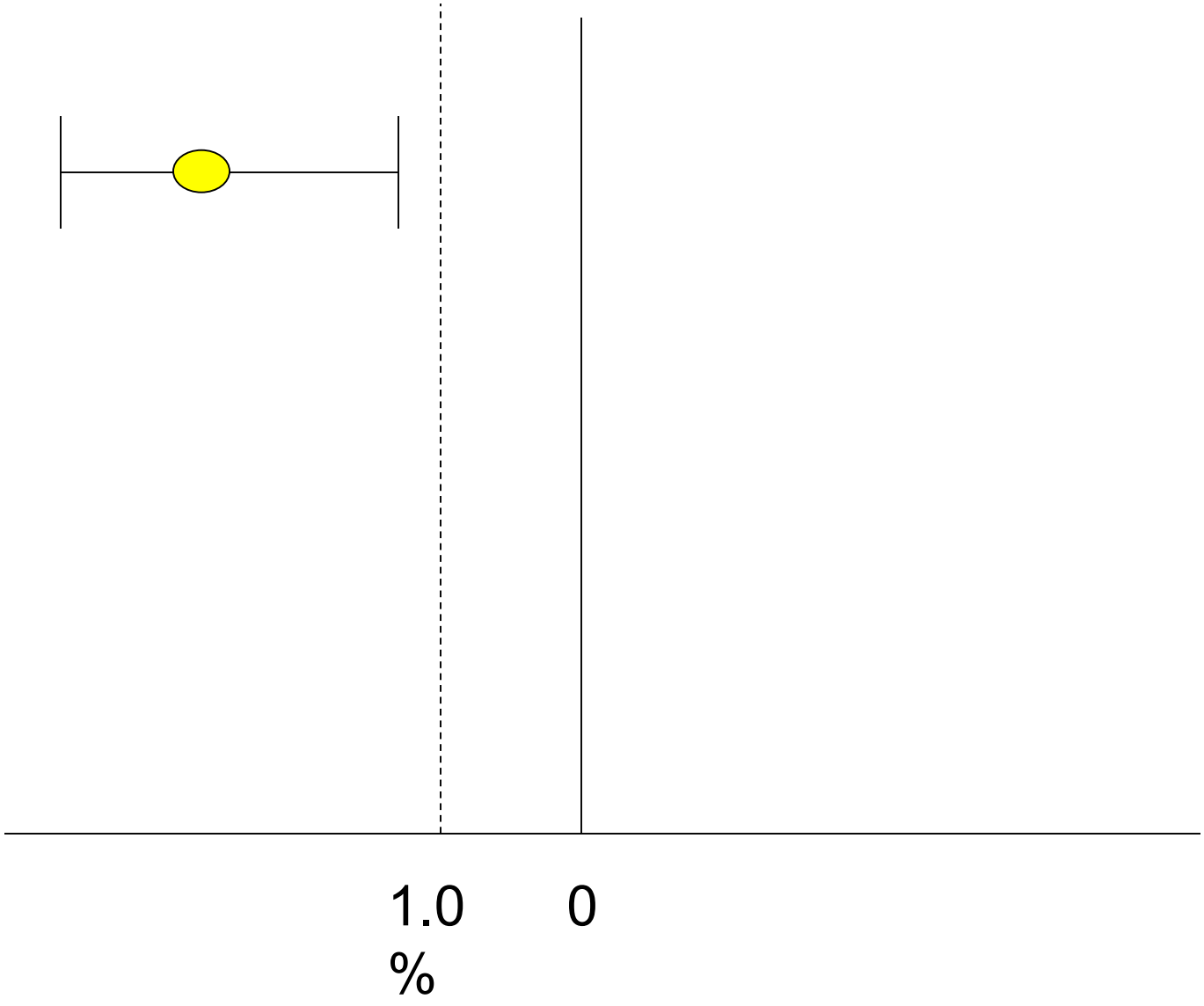


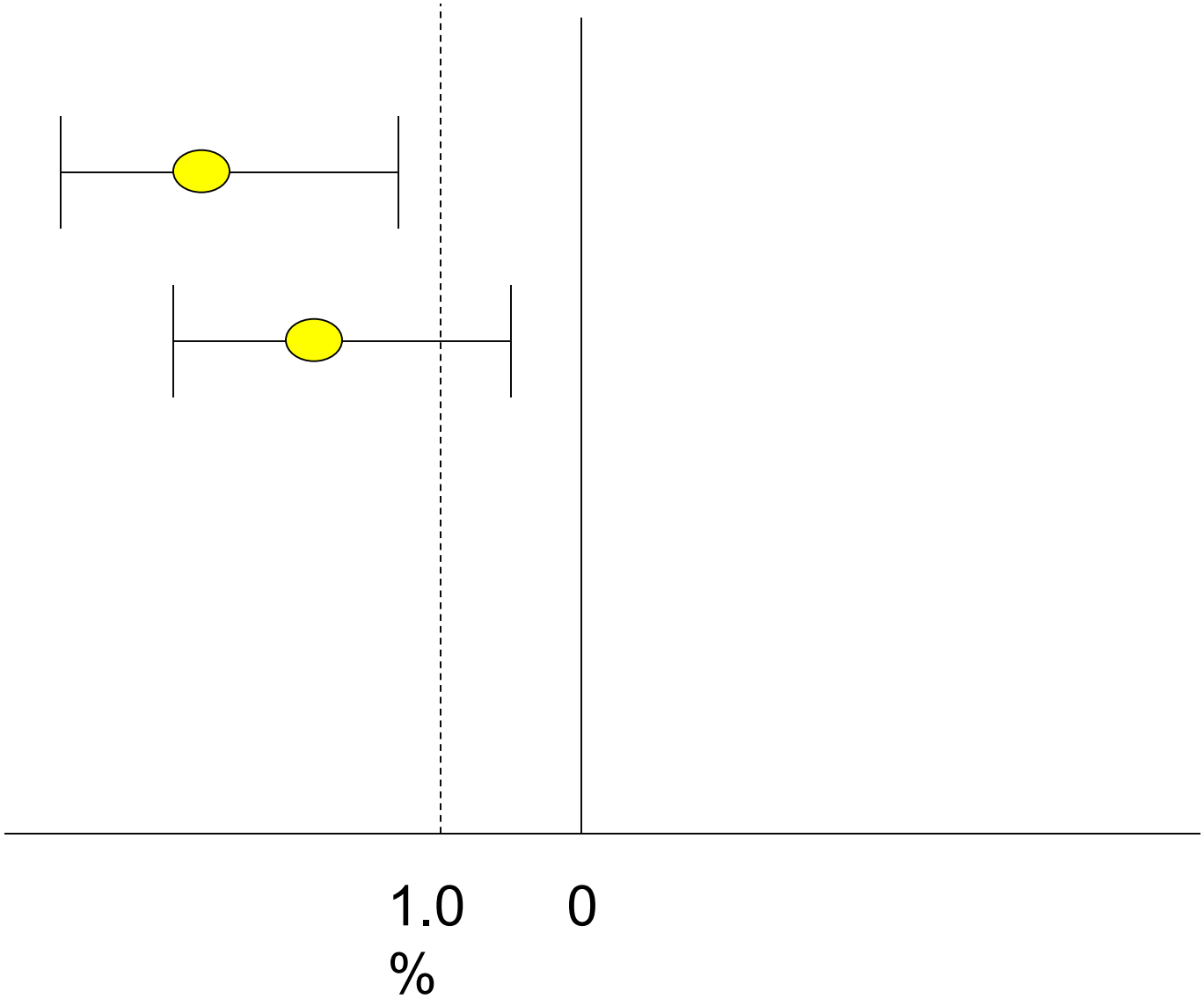
P≥0.05 potrebbe nascondere differenze clinicamente rilevanti

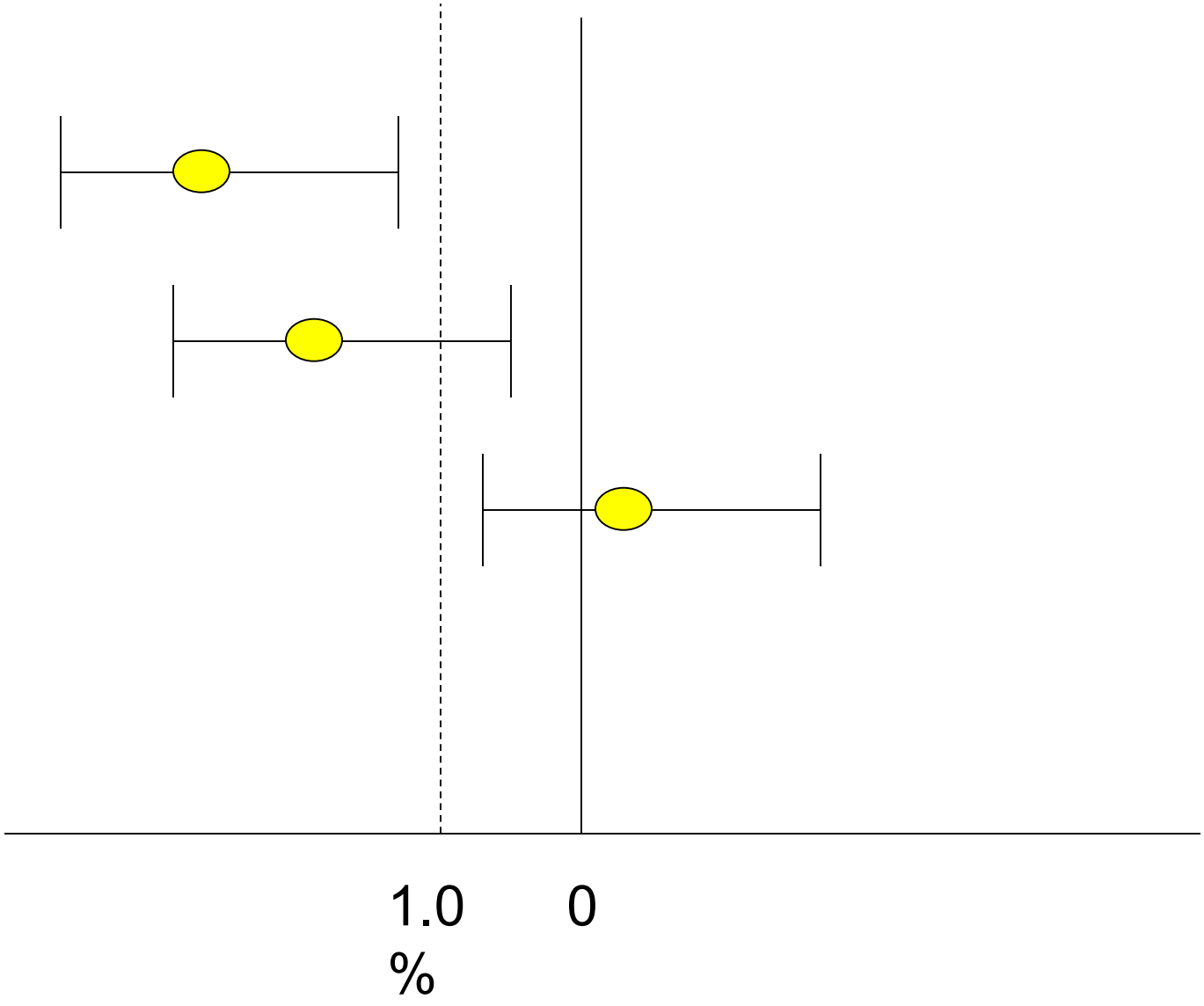


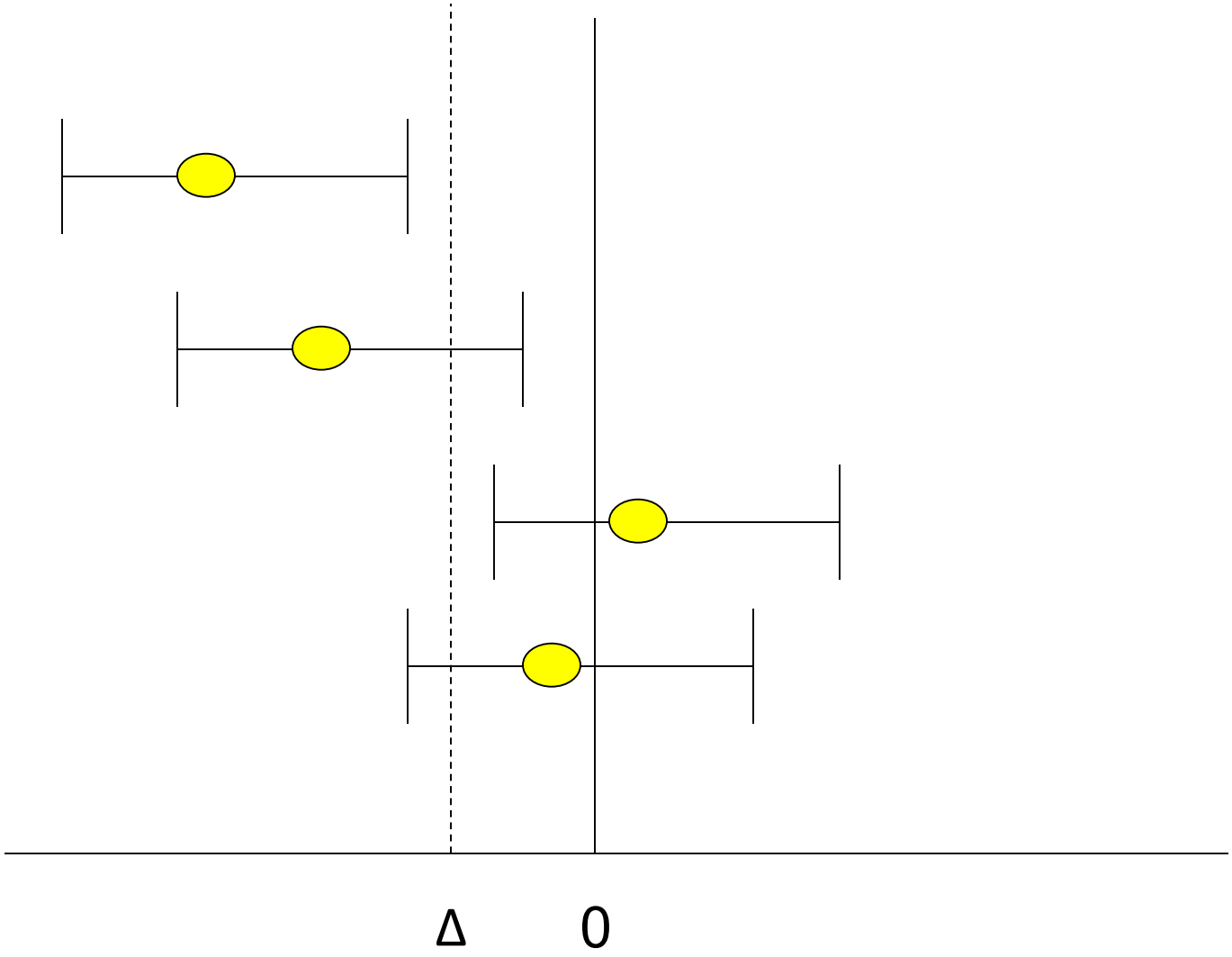
Example: clopidogrel or ASA?

- pts with threatened stroke in secondary prevention
- RCT of clopidogrel vs ASA
  - 19,185 patients
- ischaemic stroke, MI, or vascular death compared
  - 939 events (5·32%) clopidogrel
  - 1021 events (5·83%) with aspirin
- RR 0.91 (95% CI 0.83 – 0.99) (p=0·043)
- imprecision?


1.0
0
%


1.0
0
%


1.0
0
%


Δ
0

P value – Statistical significance

Confidence Interval – Clinical significance

MULTIPLICITY

Multiplicity

- Ogni volta che si esegue un test di significatività statistica, si corre il rischio di essere «ingannati» da fluttuazioni casuali che portano a concludere che si sia di fronte ad un effetto reale quando, invece, non esiste.

Multiplicity

- Questo è chiamato errore di tipo I. Quando diciamo che abbiamo bisogno di $p < 0.05$ per la significatività, diciamo che vogliamo limitare il tasso di errore di Tipo I al 5 percento. Ma il tasso di errore del 5% si applica a tutti i test statistici eseguiti.

REMEMBER

The more analyses you perform on a data set, the more your overall alpha level increases: Perform two tests and your chance of at least one of them coming out falsely significant is about 10 percent; run 40 tests, and the overall alpha level jumps to 87 percent. This is referred to as the problem of *multiplicity*, or as *Type I error inflation*.

Multiplicity

- Quando si testano diverse ipotesi, come il confronto di diverse variabili a diversi time-points o tra gruppi diversi, rimane da decidere quale tipo di strategia per il controllo di alfa si vuole implementare.

FOUR BASIC APPROACHES TO MULTIPLE COMPARISONS PROBLEMS

1. Ignore the problem; report all interesting results
2. Perform all desired tests at the nominal (typically 0.05) level and warn reader that no accounting has been taken for multiple testing
3. Limit yourself to only one test
4. Adjust the p-values/confidence interval widths in some statistically valid way

   - Divide desired α by the number of comparisons (Bonferroni)

ANALISI PER SOTTOGRUPPI

Only one thing is worse than doing subgroup analyses---believing the results

R. Peto

What are subgroup?

- An analysis of treatment effects within subgroups of patients enrolled on a clinical trial, based on baseline characteristics, who might be expected to respond to treatment differently

- “Should all patients be given XYZ? Can/should treatment be limited to a selected group?”

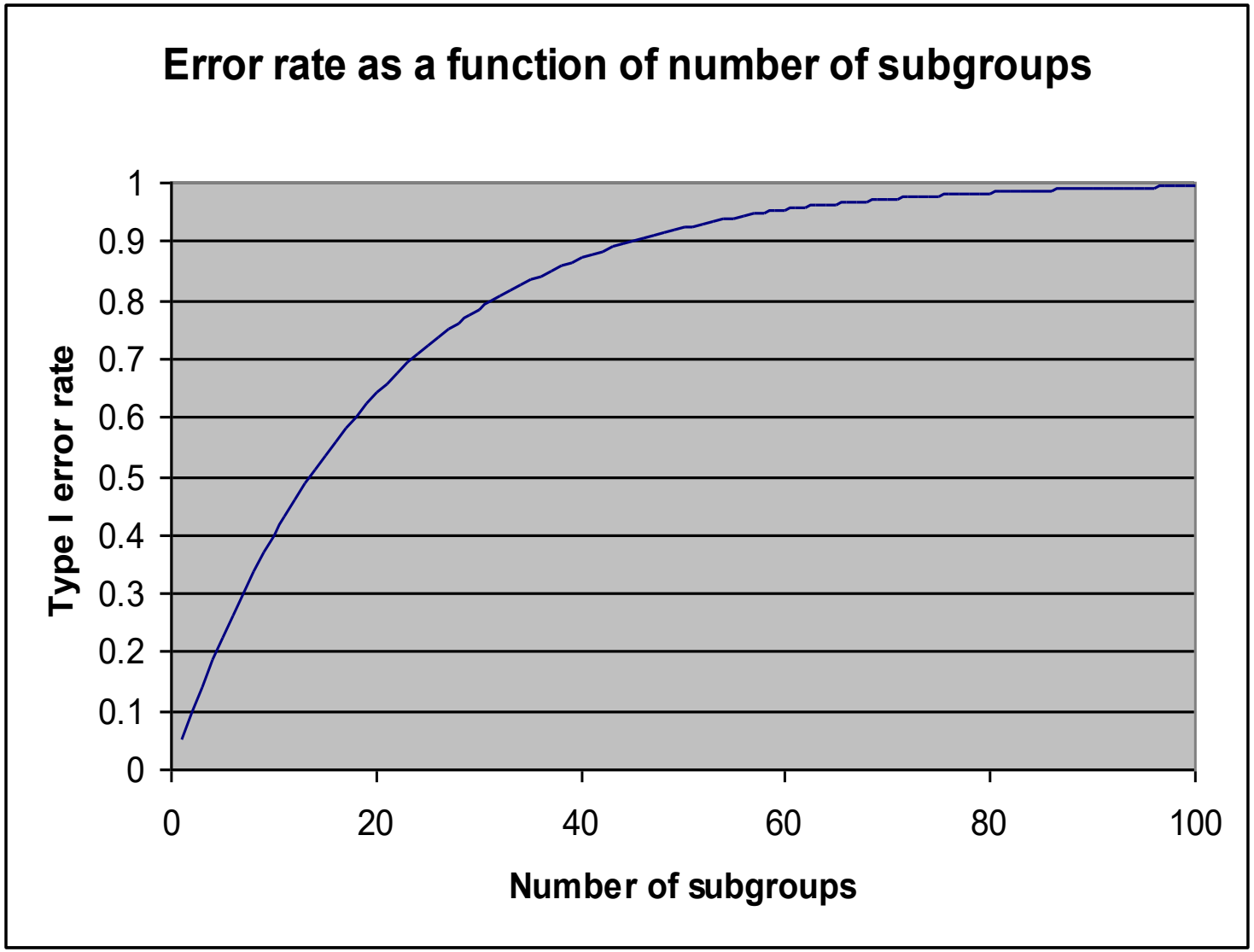
Frequency of Subgroup Analyses

- Approximately 50% of reports of randomized clinical trials contain at least one subgroup analysis (Pocock et al 1987)

General Assumptions in Subgroup Analysis

- Hypotheses tested usually address an overall or ‘average’ treatment effect in the study population
- No assumption of homogeneity of effect across subgroups - interaction
- Direction, not magnitude, of the treatment effect is expected be the same in subgroups

When multiple subgroup analyses are performed, the probability of a false positive finding can be substantial

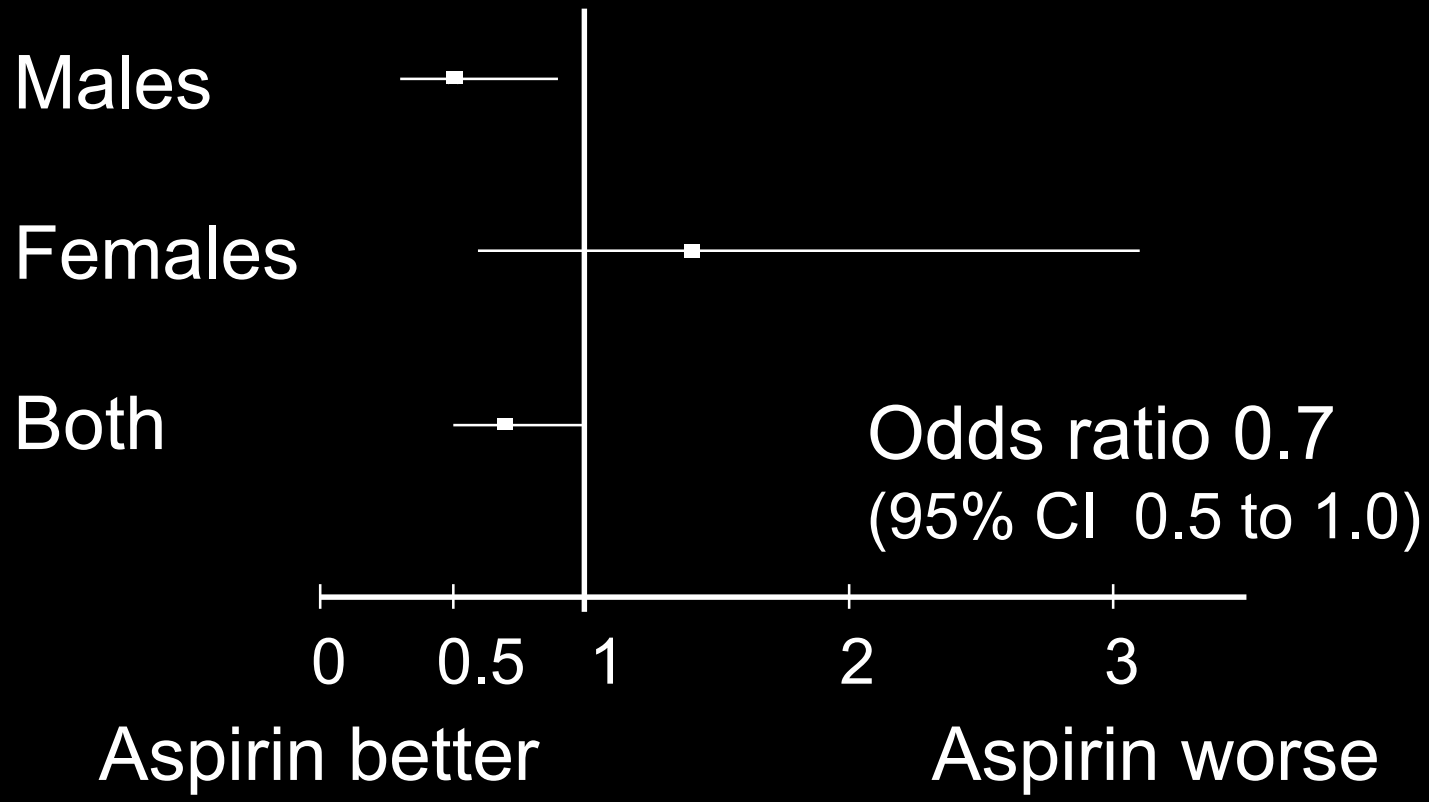

Error rate as a function of number of subgroups
Type I error rate
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
20
40
60
80
100
Number of subgroups

Post-hoc analysis

- Unplanned analyses (exploratory)
  - Analyses suggested by the data
  - Exhaustive search for differential treatment effects by subgroups (data dredging)
  - Inflated, and generally unknown, error rates

Inappropriate subgroup analysis can kill

Canadian Co-op Study Group 1978: relative odds of stroke or death in 585 TIA/stroke patients treated longterm with aspirin vs no aspirin



Gent, Circulation 1980

Impact of this result

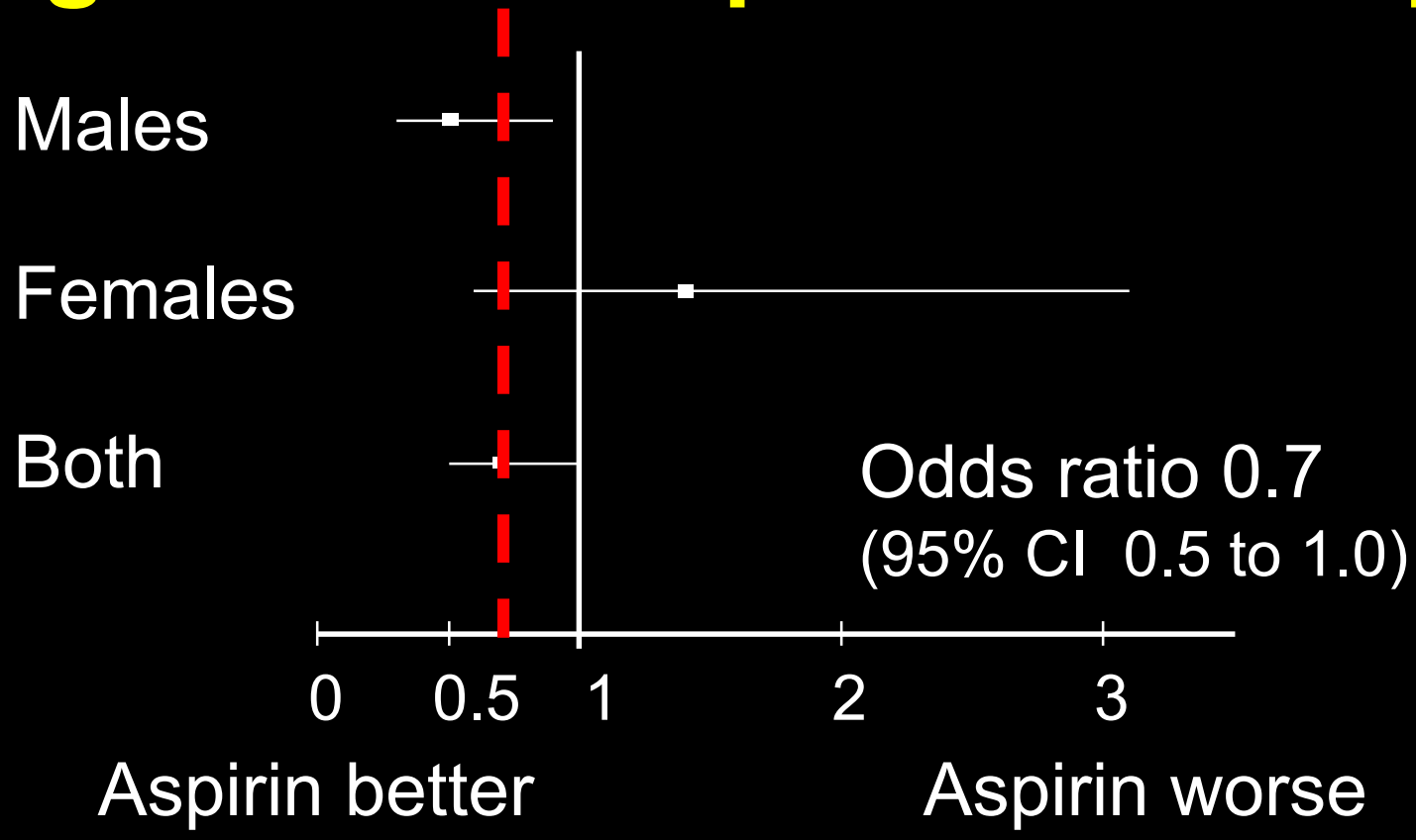
- FDA did not licence aspirin for stroke prevention in women
- Millions of women were denied effective therapy
- Many avoidable strokes and deaths from vascular disease occurred

the question is NOT: ‘Is the treatment effect in this subgroup statistically significantly different from zero?’

BUT…
are there any differences in the treatment effect *between* the various subgroups?

The correct statistical procedures are either a test of heterogeneity or a test for *interaction*

Canadian Co-op Study Group 1978: relative odds of stroke or death in 585 TIA/stroke patients treated longterm with aspirin vs no aspirin



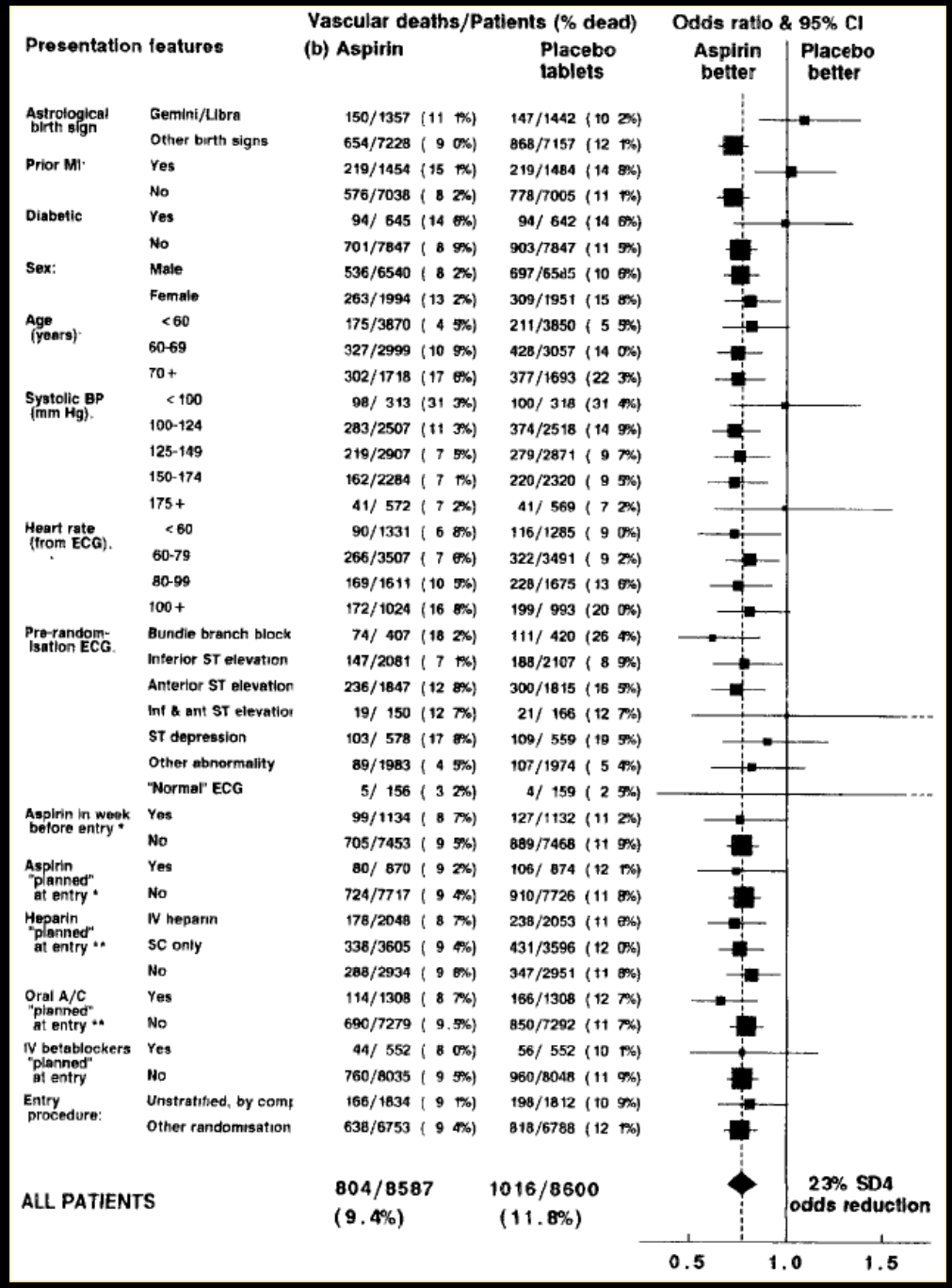
ISIS-2: aspirin vs control - effects on vascular death in 17,187 patients with acute myocardial infarction (MI)

	Aspirin	Control	Relative risk reduction
Overall trial result	9.4%	11.8%	20%

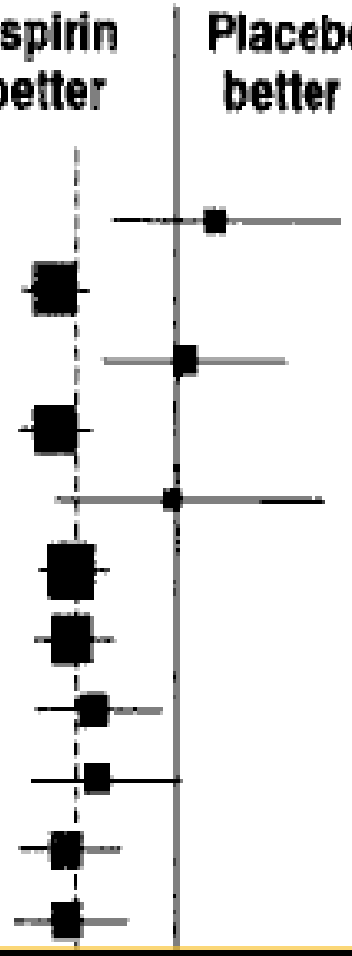
P < 0·00001

When this paper was submitted to the Lancet, the editors urged the researchers to include nearly 40 subgroup analyses.

The investigators reluctantly agreed, under the condition that they could provide a subgroup analysis of their own to illustrate their unreliability.


Presentation features
Vascular deaths/Patients (% dead)
(b) Aspirin
Placebo tablets
Odds ratio & 95% CI
Aspirin better
Placebo better
Astrological birth sign
Gemini/Libra 150/1357 (11 1%) 147/1442 (10 2%)
Other birth signs 654/7228 (9 0%) 868/7157 (12 1%)
Prior MI
Yes 219/1454 (15 1%) 219/1484 (14 8%)
No 576/7038 (8 2%) 778/7005 (11 1%)
Diabetic
Yes 94/ 645 (14 6%) 94/ 642 (14 6%)
No 701/7847 (8 9%) 903/7847 (11 5%)
Sex:
Male 536/6540 (8 2%) 697/6585 (10 6%)
Female 263/1994 (13 2%) 309/1951 (15 8%)
Age (years)
< 60 175/3870 (4 5%) 211/3850 (5 5%)
60-69 327/2999 (10 9%) 428/3057 (14 0%)
70 + 302/1718 (17 6%) 377/1693 (22 3%)
Systolic BP (mm Hg).
< 100 98/ 313 (31 3%) 100/ 318 (31 4%)
100-124 283/2507 (11 3%) 374/2518 (14 9%)
125-149 219/2907 (7 5%) 279/2871 (9 7%)
150-174 162/2284 (7 1%) 220/2320 (9 5%)
175 + 41/ 572 (7 2%) 41/ 569 (7 2%)
Heart rate (from ECG).
< 60 90/1331 (6 8%) 116/1285 (9 0%)
60-79 266/3507 (7 6%) 322/3491 (9 2%)
80-99 169/1611 (10 5%) 228/1675 (13 6%)
100 + 172/1024 (16 8%) 199/ 993 (20 0%)
Pre-random-isation ECG.
Bundle branch block 74/ 407 (18 2%) 111/ 420 (26 4%)
Inferior ST elevation 147/2081 (7 1%) 188/2107 (8 9%)
Anterior ST elevation 236/1847 (12 8%) 300/1815 (16 5%)
Inf & ant ST elevatio 19/ 150 (12 7%) 21/ 166 (12 7%)
ST depression 103/ 578 (17 8%) 109/ 559 (19 5%)
Other abnormality 89/1983 (4 5%) 107/1974 (5 4%)
"Normal" ECG 5/ 156 (3 2%) 4/ 159 (2 5%)
Aspirin in week before entry *
Yes 99/1134 (8 7%) 127/1132 (11 2%)
No 705/7453 (9 5%) 889/7468 (11 9%)
Aspirin "planned" at entry *
Yes 80/ 870 (9 2%) 106/ 874 (12 1%)
No 724/7717 (9 4%) 910/7726 (11 8%)
Heparin "planned" at entry **
IV heparin 178/2048 (8 7%) 238/2053 (11 6%)
SC only 338/3605 (9 4%) 431/3596 (12 0%)
No 288/2934 (9 8%) 347/2951 (11 8%)
Oral A/C "planned" at entry **
Yes 114/1308 (8 7%) 166/1308 (12 7%)
No 690/7279 (9.5%) 850/7292 (11 7%)
IV betablockers "planned" at entry
Yes 44/ 552 (8 0%) 56/ 552 (10 1%)
No 760/8035 (9 5%) 960/8048 (11 9%)
Entry procedure:
Unstratified, by comp 166/1834 (9 1%) 198/1812 (10 9%)
Other randomisation 638/6753 (9 4%) 818/6788 (12 1%)
ALL PATIENTS
804/8587 (9.4%)
1016/8600 (11.8%)
23% SD4 odds reduction
0.5
1.0
1.5

Presentation features		Vascular deaths/Patients (% dead) (b) Aspirin	Placebo tablets	Odds ratio & 95% CI Aspirin better	Placebo better
Astrological birth sign	Gemini/Libra	150/1357 (11 1%)	147/1442 (10 2%)		
	Other birth signs	654/7228 (9 0%)	868/7157 (12 1%)		
Prior MI	Yes	219/1454 (15 1%)	219/1484 (14 8%)		
	No	576/7038 (8 2%)	778/7005 (11 1%)		
Diabetic	Yes	94/ 645 (14 6%)	94/ 642 (14 6%)		
	No	701/7847 (8 9%)	903/7847 (11 5%)		
Sex:	Male	536/6540 (8 2%)	697/6585 (10 6%)		
	Female	263/1994 (13 2%)	309/1951 (15 8%)		
Age (years)	< 60	175/3870 (4 5%)	211/3850 (5 5%)		
	60-69	327/2999 (10 9%)	428/3057 (14 0%)		
	70 +	302/1718 (17 6%)	377/1693 (22 3%)		



Author's conclusions

Apparent harm in patients born under star sign of libra or gemini, with prior MI and diabetics, all most likely due to the play of chance

"All these subgroup analyses should, perhaps, be taken less as evidence about who benefits, than as evidence that such analyses are potentially misleading."

Pre-specified Subgroup Analyses

- Pre-specified analyses (hypothesis driven)
  - Subgroup hypotheses specified in advance in the study protocol
  - Control of error rates can, in principle, be addressed (statistics) - not always done

Pre-planned Subgroup Analyses

- Pre-planned analyses (hypothesis driven)
  - Subgroup hypotheses specified in advance
  - Control of error rates addressed (statistical analysis)

Control of Error Rates in Subgroup Analyses

- For planned subgroup analyses, the overall type I error rate can be controlled. One conservative way is to use $\alpha^* = \alpha/k$ in each of the subgroup analyses
- In this case, the power (probability of detecting real differences when present) is sharply reduced in individual subgroups
- For unplanned subgroup analyses, k is unknown so the error rates are unknown

Error Rates in Subgroup Analyses

With k independent subgroups and no difference in treatments, the probability

of at least one ‘significant’ subgroup is:

$$1- (1- \alpha)^{k}$$

For example, $\alpha = 0.05$, $k = 10$ yields

$$1- (1- 0.05)^{10} = 0.40$$

Conclusioni

- Analisi pre-pianificata di sottogruppi

DIMOSTRATIVA

- Analisi pre-specificata di sottogruppi

DUBBIA

- Analisi post-hoc di sottogruppi

SUGGESTIVA