Subgroup Analysis

Naitee Ting

Boehringer-Ingelheim

All points in the slides represent my personal point of view. They do not necessarily reflect the position of Boehringer-Ingelheim

Emerging Topics in Statistics and Biostatistics

Naitee Ting · Joseph Cappelleri · Shuyen Ho · (Din) Ding-Geng Chen *Editors*

Design and Analysis of Subgroups with Biopharmaceutical Applications

This book provides an overview of the theories and applications on subgroups in the biopharmaceutical industry. Drawing from a range of expert perspectives in academia and industry, this collection offers an overarching dialogue about recent advances in biopharmaceutical applications, novel statistical and methodological developments, and potential future directions.

The volume covers topics in subgroups in clinical trial design; subgroup identification and personalized medicine; and general issues in subgroup analyses, including regulatory ones. Included chapters present current methods, theories, and case applications in the diverse field of subgroup application and analysis. Offering timely perspectives from a range of authoritative sources, the volume is designed to have wide appeal to professionals in the pharmaceutical industry and to graduate students and researchers in academe and government.

ISBN 978-3-030-40104-7

9 783030 401047

▸ springer.com

Ting · Cappelleri · Ho · Chen *Eds.*

Design and Analysis of Subgroups with Biopharmaceutical Applications

Emerging Topics in Statistics and Biostatistics

Naitee Ting
Joseph Cappelleri
Shuyen Ho
(Din) Ding-Geng Chen *Editors*

Design and Analysis of Subgroups with Biopharmaceutical Applications

Springer

Development of an RA drug

- The test drug was developed to treat rheumatoid arthritis (RA) and osteoarthritis (OA)
- In the 1980’s two classes of RA drugs were available
  - Non-Steroidal Anti-Inflammatory Drugs (NSAID)
  - Disease Modifying Anti-Rheumatic Drugs (DMARD)

Development of an RA drug

- NSAID’s are quick acting, symptom control, and more on pain relief
- DMARD’s are slow acting (several months), and help slow down the disease progression

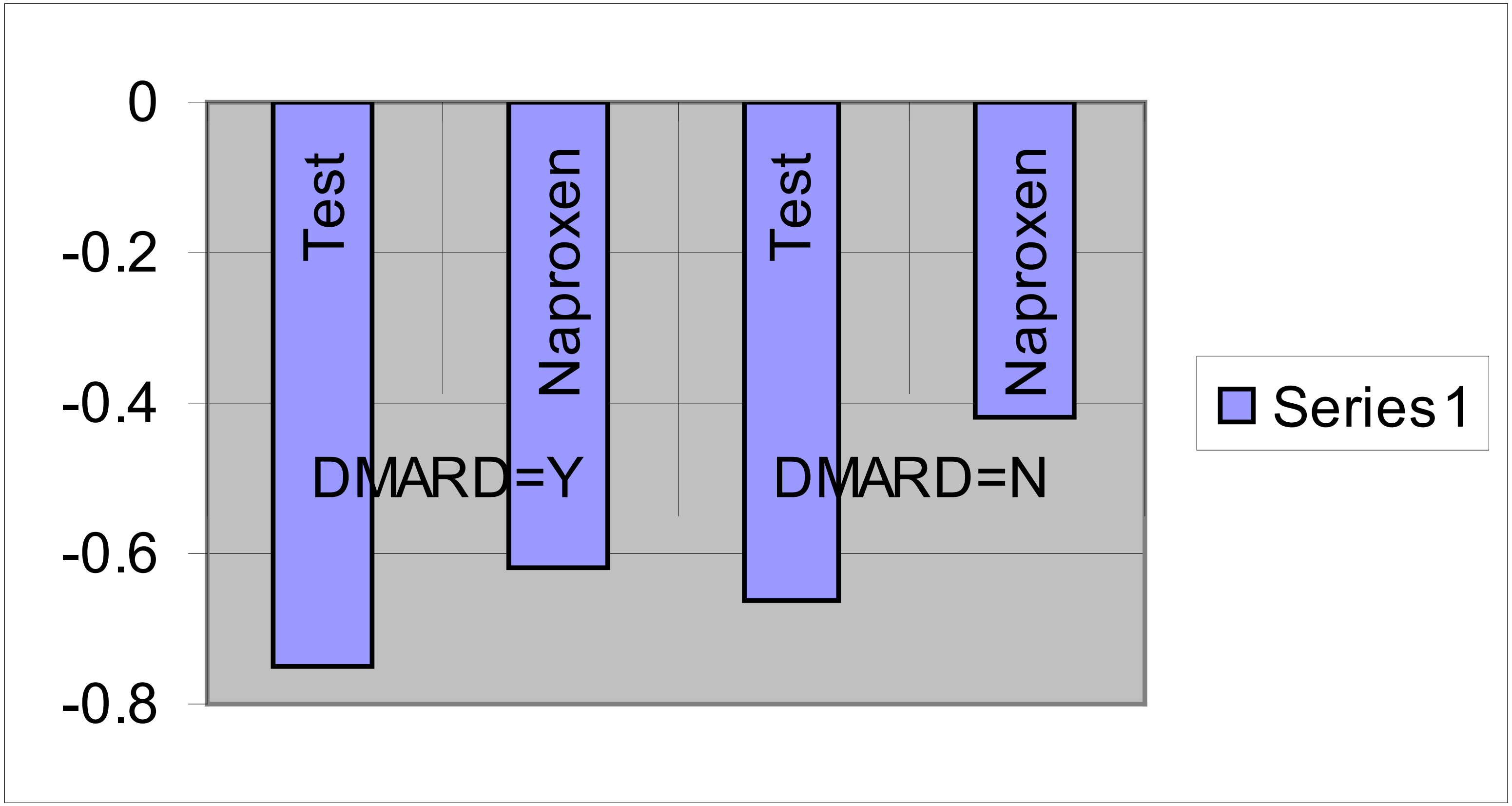
Development of an RA drug

- The test drug was developed as an NSAID
- A long term protocol compares Test and Naproxen (an NSAID) in treatment of RA patients
- Five year study, with primary time point at 6 months
- Randomized with stratification on DMARD usage

Development of an RA drug

- Primary endpoint is physician assessment of disease activity
  - No symptom (1), Mild (2), Moderate (3), Severe (4) and Very severe (5)
- Test drug -0.76, Naproxen -0.57, p=0.0103
- Subgroup analysis – by DMARD usage

Physician Assessment



Development of an RA drug

- Test drug has better efficacy than Naproxen
- Results from subgroup analysis indicate Test drug may be a DMARD
- Further development demonstrate DMARD efficacy of the Test drug
- Changed the direction of development based on subgroup findings

Intent-to-treat principle

- Analyze as randomized
- Include all subjects
- Imagine a wonder drug cures 9999 patients out of 10,000
- One died
- Can we exclude the death as an outlier?
- Report all data – as randomized

Background

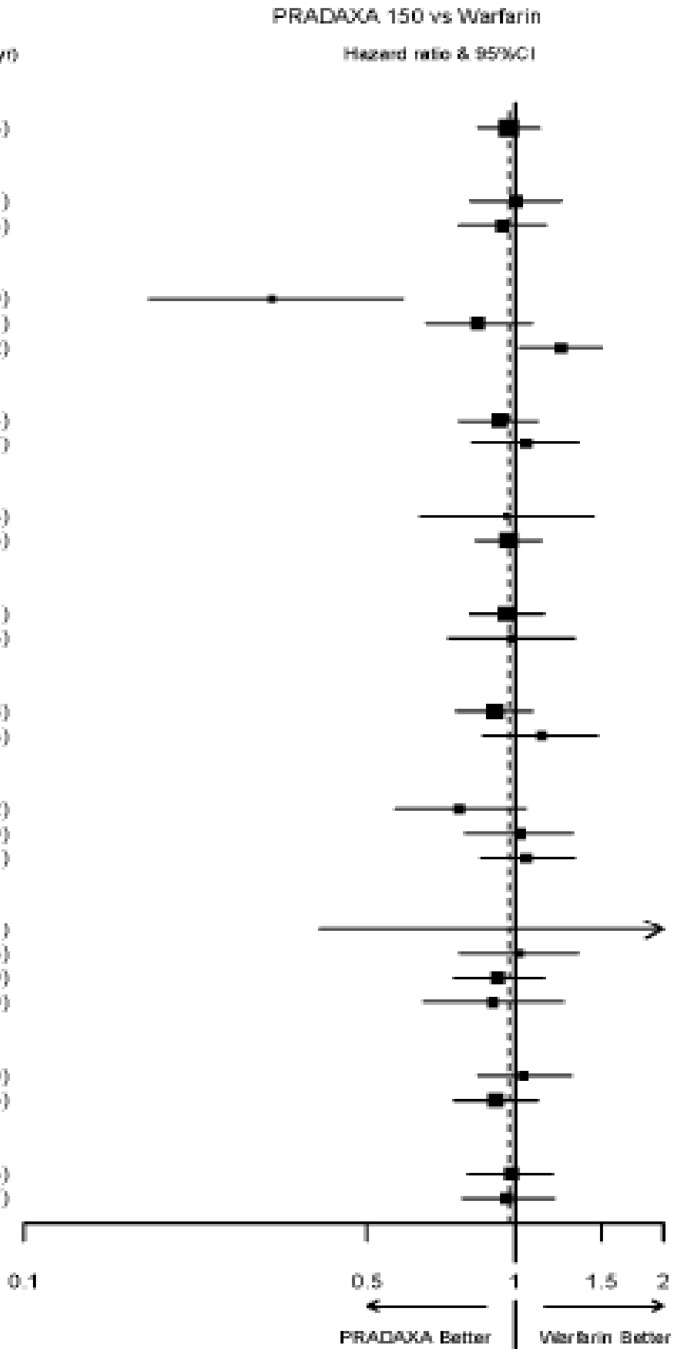
- The primary analysis for a study is typically the ITT (or the Full Analysis Set), or the PP (Per Protocol) Set
- Any analysis based on part of the entire analysis set is considered a subgroup analysis
- For example, analysis by gender, by age, or by center

Background

- Reasons of subgroup analysis
  - To address particular concerns in some specific subgroups
  - To explore whether the test drug is more efficacious or more harmful in a subset
  - To provide supportive evidence to main findings
  - To generate new hypothesis of drug effect
  - To address regulatory queries

Major bleeding events, on treatment + 2 days, safety set

Subgroup	Patients	PRADAXA 150		Warfarin		HR (95%CI)
	Total no.	n	N (% per yr)	n	N(% per yr)	
All patients	18040	350	6059 (3.47)	374	5998 (3.58)	0.97 (0.84, 1.12)
VKA use at entry						
Naive (50.4%)	9091	167	3019 (3.51)	175	3082 (3.51)	1.00 (0.81, 1.24)
Experienced (49.6%)	8946	183	3039 (3.43)	199	2916 (3.64)	0.94 (0.77, 1.15)
Age (years)						
< 65 (16.5%)	2971	14	1028 (0.77)	40	950 (2.39)	0.32 (0.18, 0.59)
≥65 and < 75 (43.6%)	7864	117	2574 (2.82)	146	2635 (3.11)	0.84 (0.66, 1.06)
≥75 (39.9%)	7205	219	2457 (5.75)	188	2413 (4.62)	1.24 (1.02, 1.50)
Gender						
Male (63.6%)	11480	221	3831 (3.37)	246	3796 (3.64)	0.93 (0.77, 1.11)
Female (36.4%)	6560	129	2228 (3.65)	128	2202 (3.47)	1.05 (0.82, 1.34)
Weight (kg)						
≤60 (10.9%)	1959	43	646 (4.59)	50	683 (4.78)	0.96 (0.64, 1.44)
> 60 (89.1%)	16074	307	5412 (3.35)	324	5312 (3.45)	0.97 (0.83, 1.13)
History of stroke/TIA						
No (80.0%)	14428	264	4827 (3.28)	285	4808 (3.41)	0.96 (0.81, 1.14)
Yes (20.0%)	3612	86	1232 (4.20)	89	1190 (4.28)	0.98 (0.73, 1.32)
Diabetes at baseline						
No (76.7%)	13836	239	4661 (3.06)	271	4593 (3.36)	0.91 (0.76, 1.08)
Yes (23.3%)	4204	111	1398 (4.87)	103	1405 (4.33)	1.13 (0.86, 1.47)
CHADS2 score						
≤1 (31.9%)	5763	72	1955 (2.10)	91	1860 (2.72)	0.77 (0.57, 1.05)
= 2 (35.6%)	6422	119	2129 (3.37)	127	2212 (3.29)	1.02 (0.79, 1.31)
≥3 (32.5%)	5855	159	1975 (5.06)	156	1926 (4.81)	1.05 (0.85, 1.32)
CrCL (mL/min)						
< 30 (0.4%)	74	3	31 (10.28)	1	29 (2.57)	3.84 (0.40,36.90)
≥30 and ≤50 (18.4%)	3327	105	1152 (6.18)	101	1048 (6.05)	1.02 (0.77, 1.34)
> 50 and ≤80 (45.8%)	8269	161	2770 (3.51)	184	2794 (3.80)	0.92 (0.75, 1.14)
> 80 (31.3%)	5641	70	1880 (2.07)	79	1872 (2.29)	0.90 (0.65, 1.25)
Region						
USA (29.7%)	5352	162	1811 (5.23)	161	1774 (5.00)	1.04 (0.84, 1.30)
OUS (70.3%)	12688	188	4248 (2.89)	213	4224 (2.95)	0.91 (0.75, 1.11)
ASA use at baseline						
No (60.3%)	10887	195	3721 (3.08)	202	3587 (3.15)	0.98 (0.80, 1.19)
Yes (39.7%)	7153	155	2338 (4.12)	172	2431 (4.27)	0.96 (0.78, 1.20)



Guidelines

- Increased attention paid to issue by journals, regulators, etc.
  - NEJM guidelines (Nov, 2007)

"Investigators frequently use analyses of subgroups of study participants to extract as much information as possible. Such analyses, …, may provide useful information for the care of patients and for future research. However, subgroup analyses also introduce analytic challenges and can lead to overstated and misleading results."

Guidelines

- ICH E-9

“When exploratory, these analyses should be interpreted cautiously; any conclusion of treatment efficacy (or lack thereof) or safety based solely on exploratory subgroup analyses are unlikely to be accepted.”

(ICH Harmonised Tripartite Guideline, 1998)

Confirm or explore?

- Hypothesis testing vs hypothesis generating
- Inferential statistics vs descriptive statistics
- Regulatory discussion/label language? Or scientific interest?
- Designed feature vs post hoc

Confirm or explore?

- Learning is an exploratory practice
- Or is it data dredging?
- Be clear that this is not confirmatory
- If it turned out that only one of the subgroups demonstrate efficacy, the regulatory agency may restrict the label

Purposes of Post-hoc Analyses Regulatory requirement

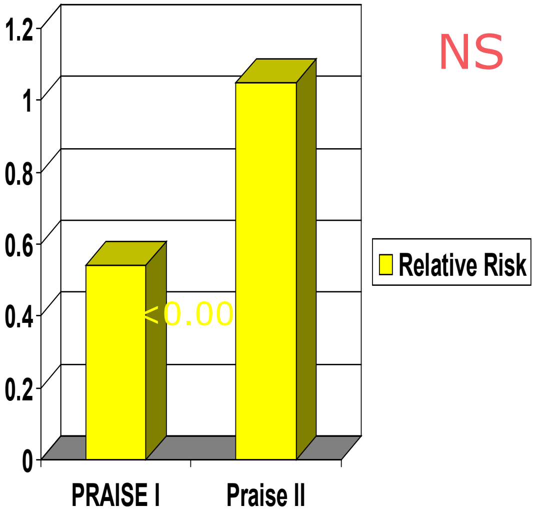
- Overall treatment effect significant in complete study:
  - Confirm consistency across *major* subgroups.
- Identify safety problems that are limited to a subgroup of patients

Pitfalls of Post-hoc Analyses

- Post-hoc vs. Pre-specified Analyses
  - Post-hoc analyses:
    - Often unclear how many were undertaken
    - Whether some were motivated by inspection of data.
  - *"… a subgroup is respectable and worthwhile when established a priori from pathophysiological principles."*

Example (False Positive)

- Praise I RCT (NEJM, 1996), Amlodipine vs. placebo in 1153 patients
- Predefined stratification (ischemic, non-ischemic)
- Mortality (non-ischemic patients) p < 0.001.
- Study repeated in Praise II enrolling non-ischemic patients.



NS

Pitfalls of Post-hoc Analyses

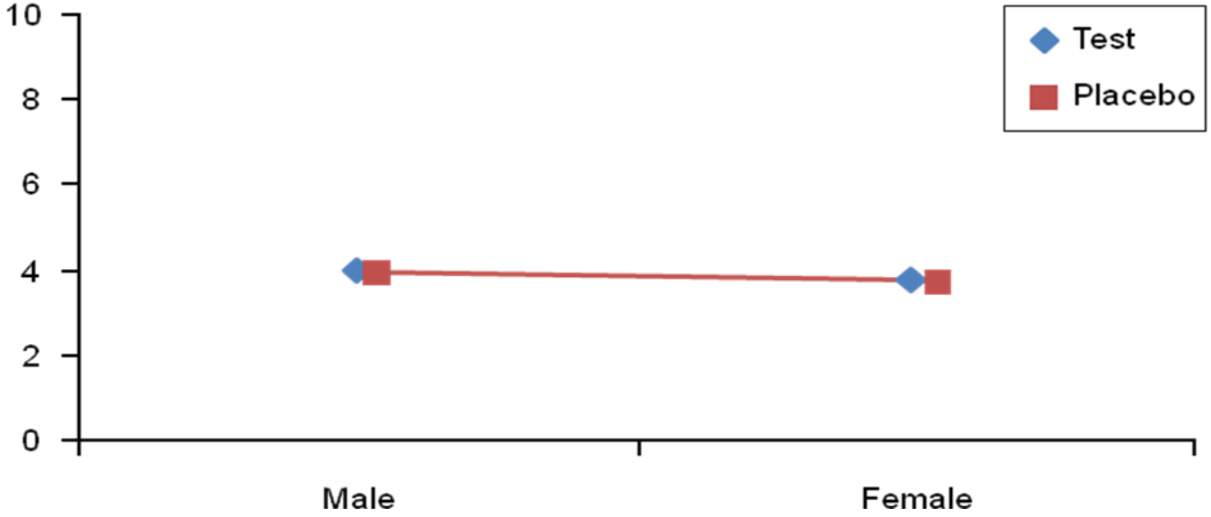
- Lack of Power: Chance of missing significant effect in a subgroup a function of size
  - Aspirin is ineffective in secondary prevention of stroke in women (Stroke, 1977, 301-14; NEJM, 1978, 53-59)
    - Refuted (BMJ, 1994, 81-106)
  - Antihypertensive treatment for primary prevention is ineffective in women (BMJ, 1985, 97-104; Ann Intern Med, 1991, 287-93)
    - Refuted (Ann Intern Med, 1997, 761-67)
  - Statin therapy is ineffective in reducing risk of coronary events in women and elderly (JAMA, 1998, 1615-22; Lancet, 2001, 351-55)
    - Refuted (Lancet, 2002, 7-22)

Treatment by Factor Interaction

- Do treatment effect differ at each level of the factor?
- For example, is the test drug better in male and worse in female?
- If test drug is better than placebo in both male and female – no interaction
- Regions, geographic areas, or other stratification factors

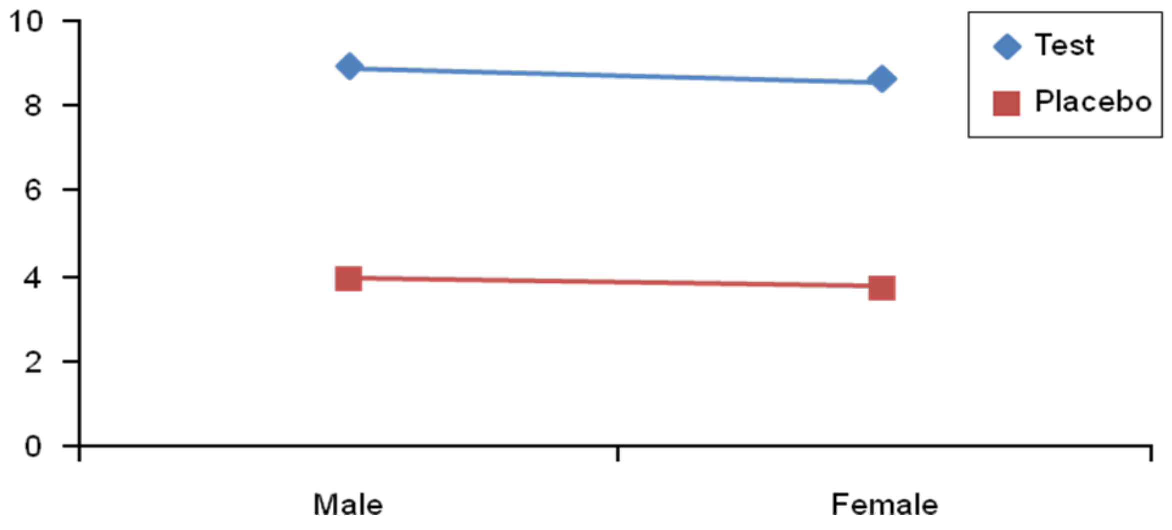
What is a treatment by sex interaction?

- When there is no treatment effect, no sex effect



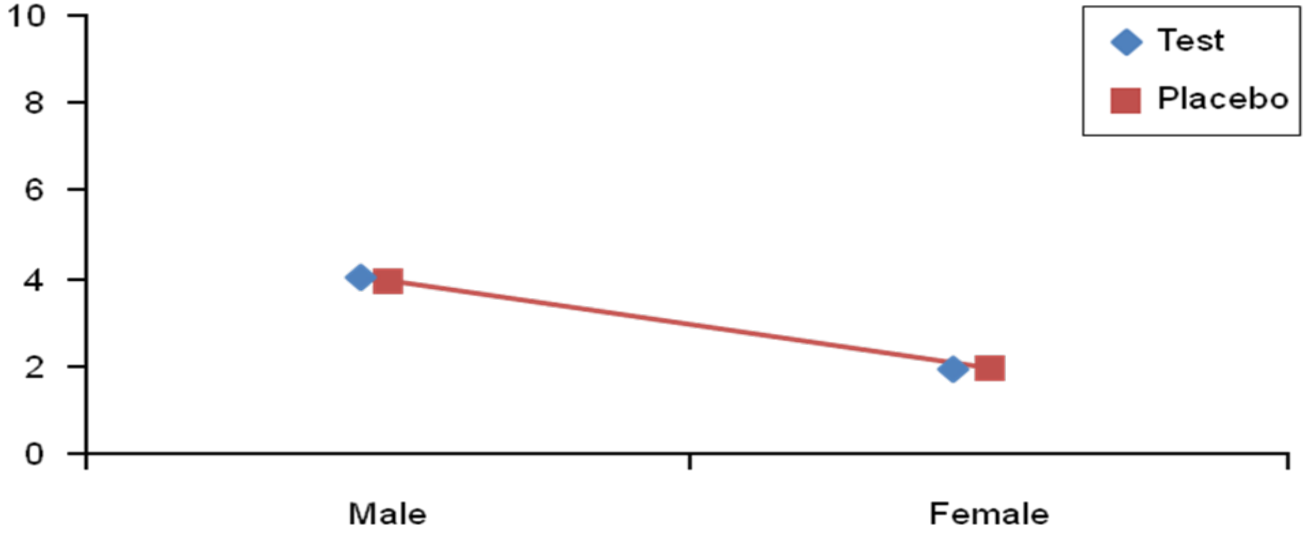
What is a treatment by sex interaction?

- When there is treatment effect, but no sex effect



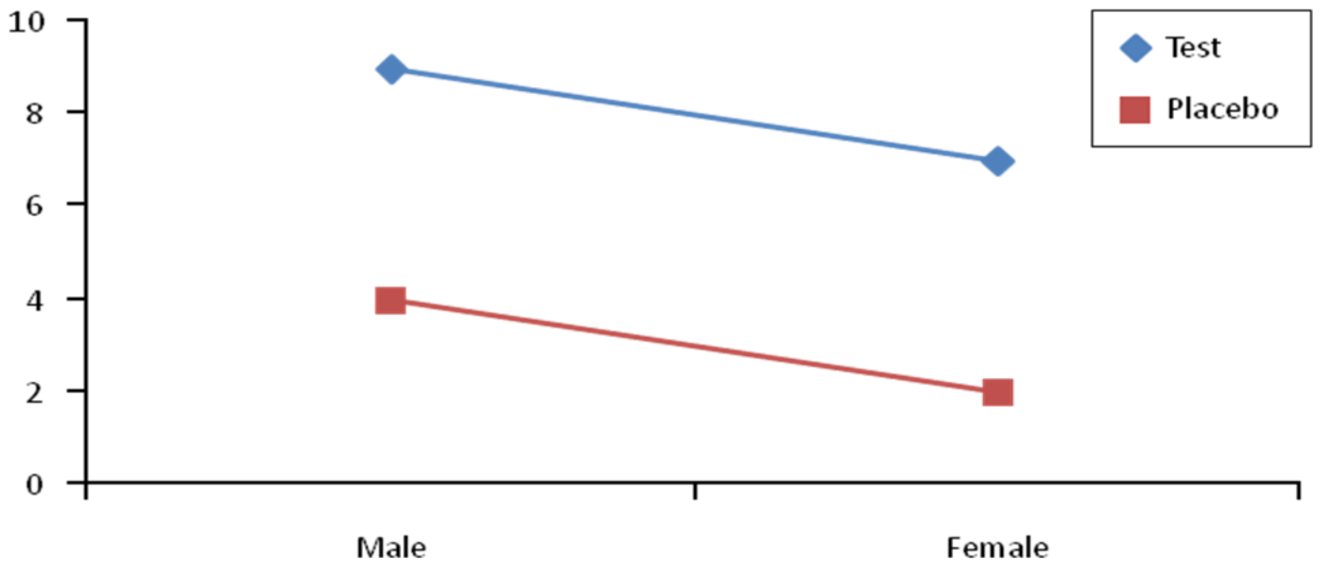
What is a treatment by sex interaction?

- When there is no treatment effect, but with sex effect



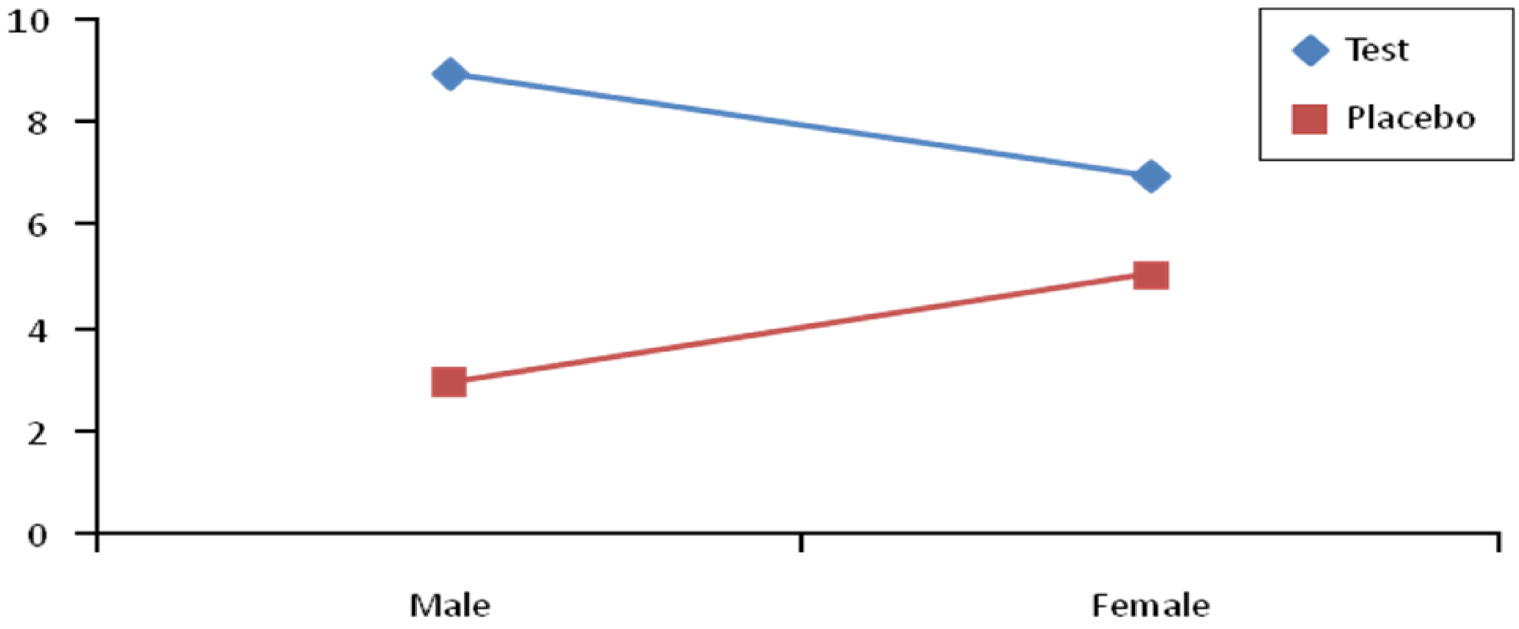
What is a treatment by sex interaction?

- When there is treatment effect, and there is sex effect – no interaction



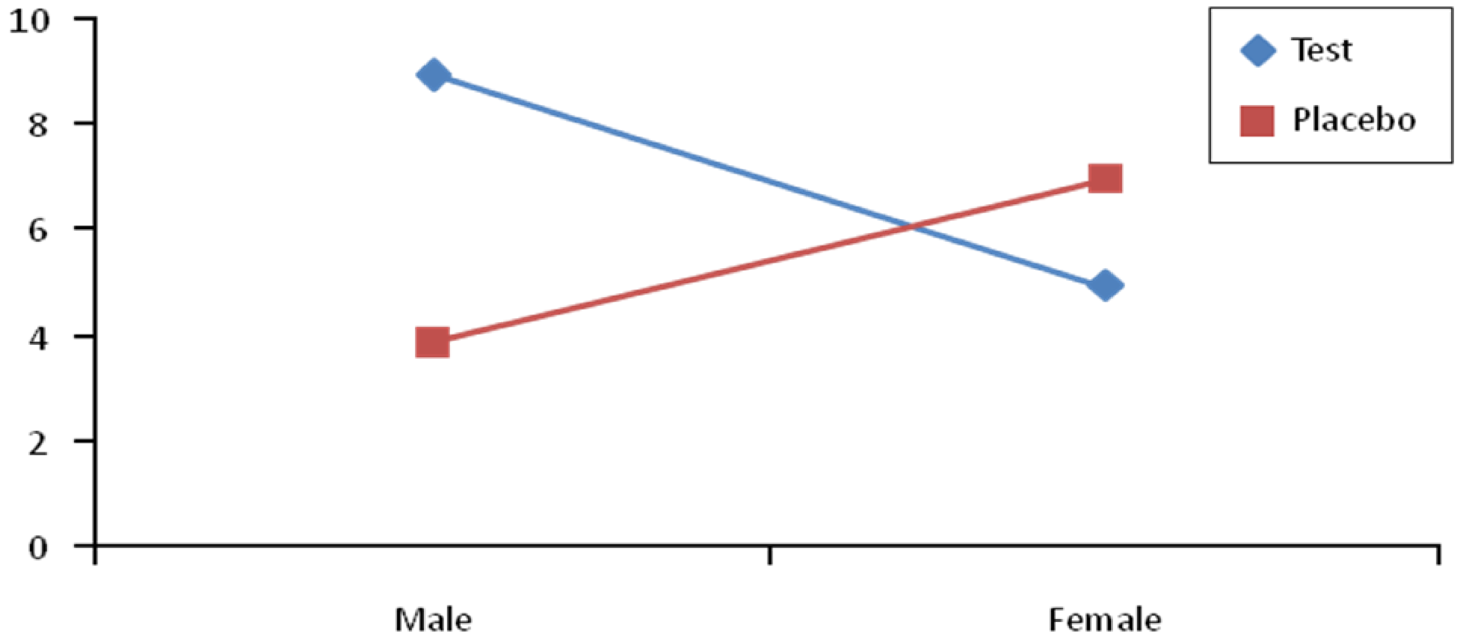
What is a treatment by sex interaction?

- When there is quantitative interaction



What is a treatment by sex interaction?

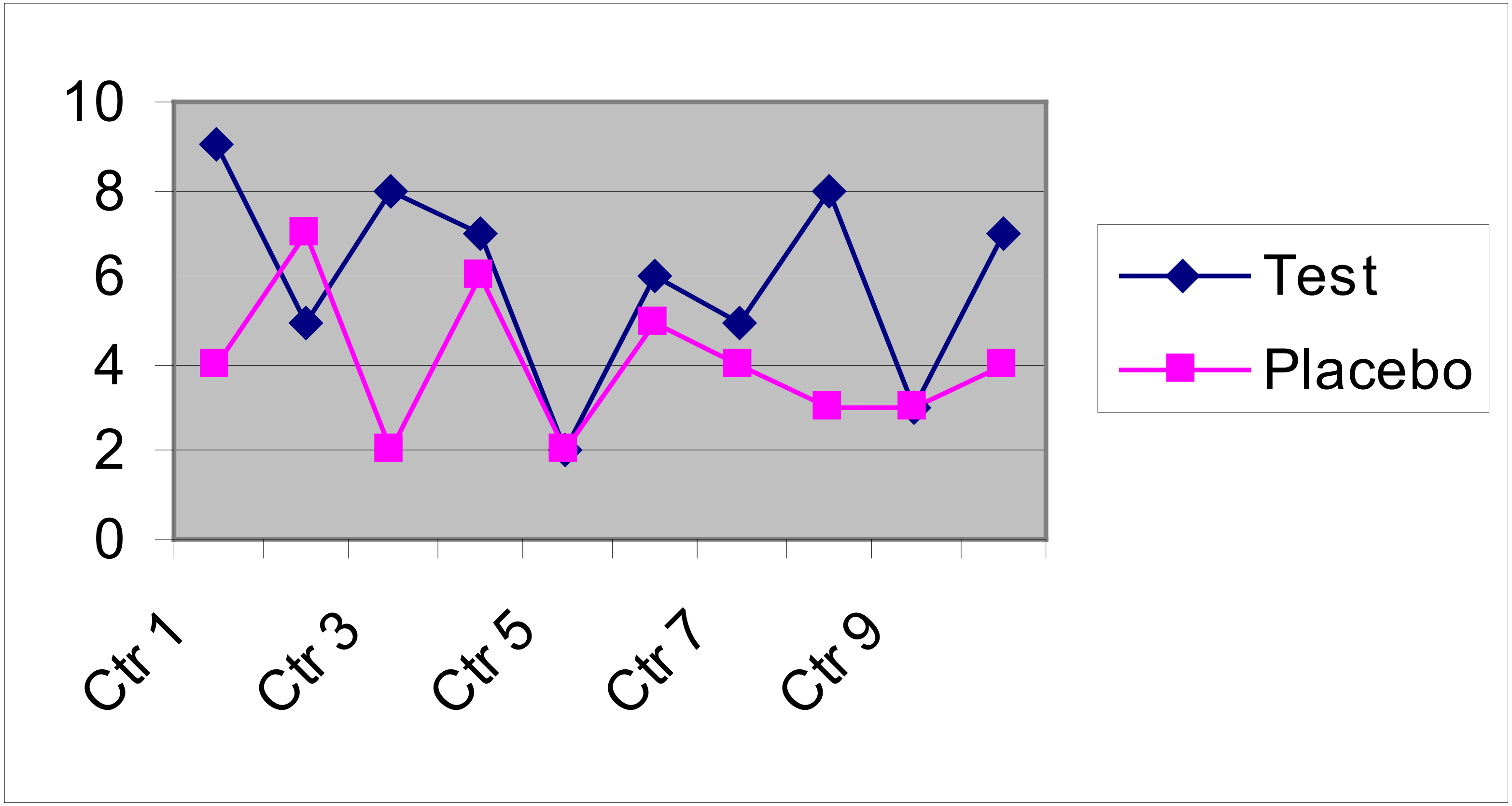
- When there is qualitative interaction



Treatment by Factor Interaction

- Another factor is center
- By center analysis is commonly applied
- Treatment by center interaction typically not significant for an active treatment
- Descriptive statistics

Treatment by center interaction



Power

- The trial is powered to study the entire study group
- There is insufficient power to make treatment comparisons in any subgroup
- Computing p-values within subgroup is not appropriate
- If there is a signal from subgroup analysis, design a new study to test this hypothesis

Power

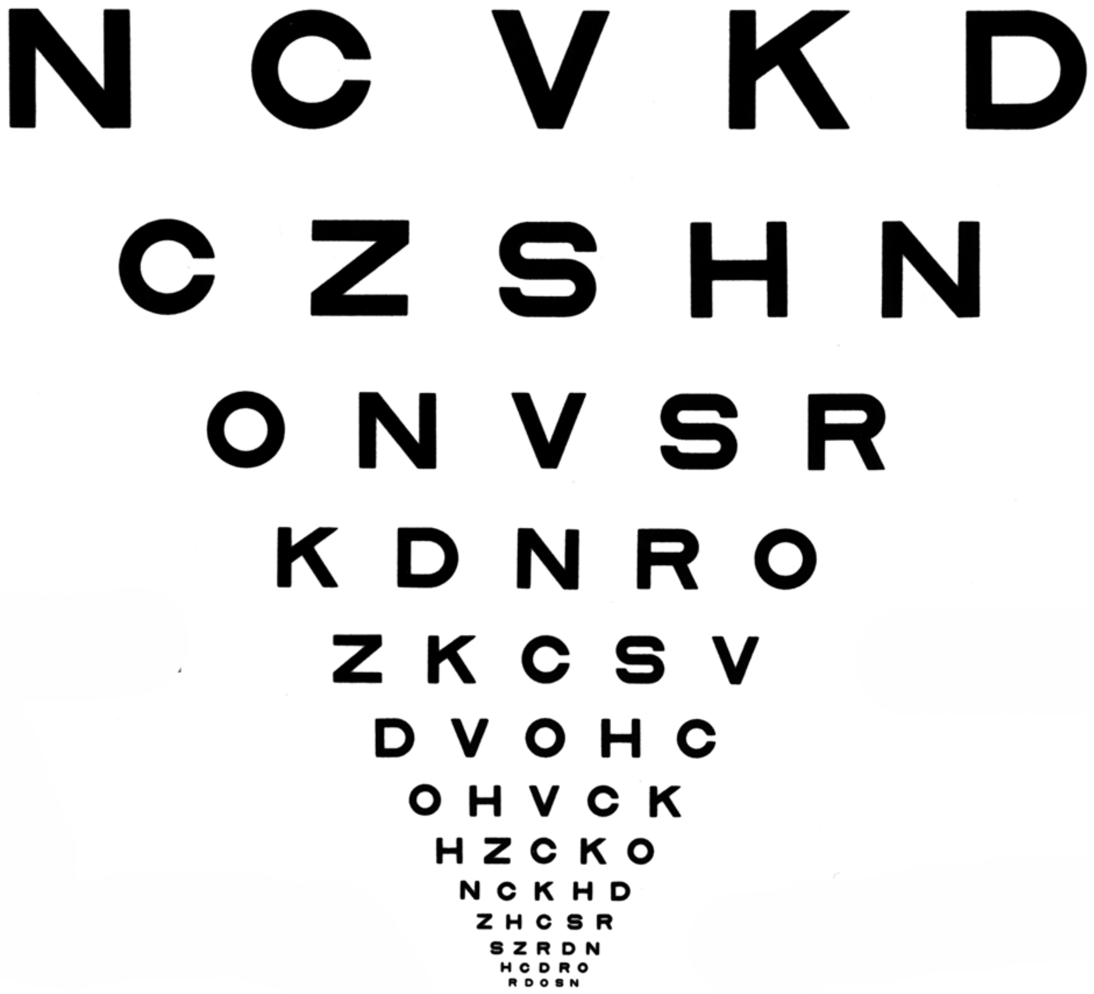
- In certain situations, the primary interest is for a specific subgroup at the design stage
- The study should be powered for that subgroup
- Stratification is recommended
- Pre-specify the primary analysis is based on the subgroup

Subgroup need to be defined at Baseline

- Post baseline characteristics may be affected by treatment
- At baseline: trt by factor -> response
- Post baselie: trt -> factor -> response (treatment effect not interpretable)
- Analysis of compliance data – may not be appropriate
- Analyze as randomized

Macugen Example

- Macugen was developed to treat age-related macular degeneration (AMD)
- Primary endpoint is change in visual acuity at 54 week post baseline
- Losing 15 letters or more is considered a non-responder
- PDT was the only available treatment during macugen development



NCVKD
CZSHN
ONVSR
KDNRO
ZKCSV
DVOHC
OHVCK
HZCKO
NCKHD
ZHCSR
SZRDN
HCDRO
RDOSN

Macugen Example

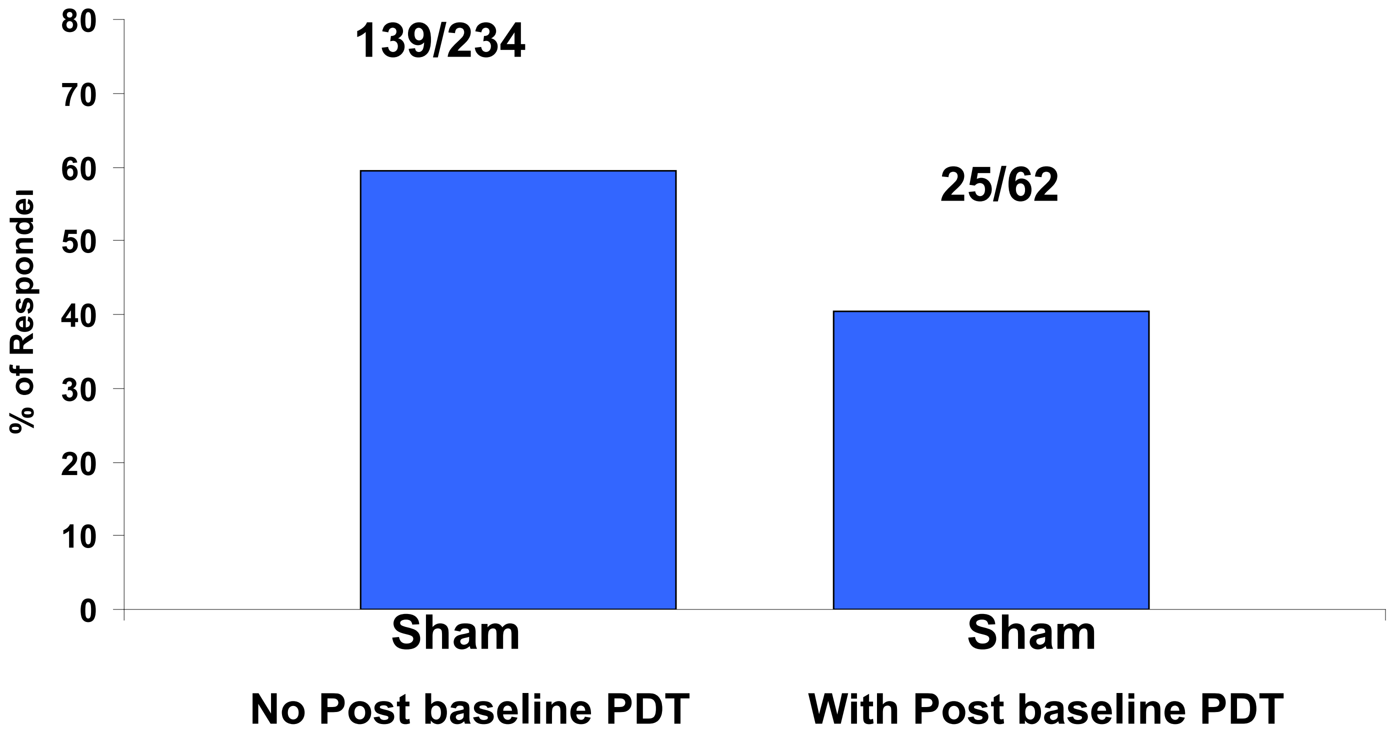
- Overall PDT usage is low
- For prior PDT usage or baseline PDT usage, there is no treatment by PDT interaction
- For post-baseline PDT, Macugen treated patients did not receive more PDT than sham treated patients
- Macugen treatment effect can not be explained by overuse of PDT

Macugen Example

- Estimates of treatment response in the presence or absence of post-baseline use of PDT is biased.
- For example, Sham treated patients with post-baseline PDT responded less well than those without PDT

Demonstrating Potential Bias

% of Sham Responders with or without post-baseline PDT, Pivotal trials



Combined analysis

- Subgroup analysis may be performed in ISS or ISE
- After pooling across studies, there may be sufficient sample size within subgroups of interest
- e.g., gender, age, race
- Or other baseline characteristics

Combined analysis

- In combined analysis, subgroup results can also be presented by study to see the consistency across studies
- Subgroup findings may be supported by the combined results
- If a particular subgroup is of interest before designing all relevant studies, it helps to pre-specify for each study, and how to combine them

Conclusion

- A clinical trial is designed to study the entire patient population. Subgroup analysis are mainly exploratory
- If there is a need, it is desirable to pre-specify at the design stage
- Results obtained from subgroups may not necessarily be replicated
- Potential consequence of restricted labeling

BACK UP

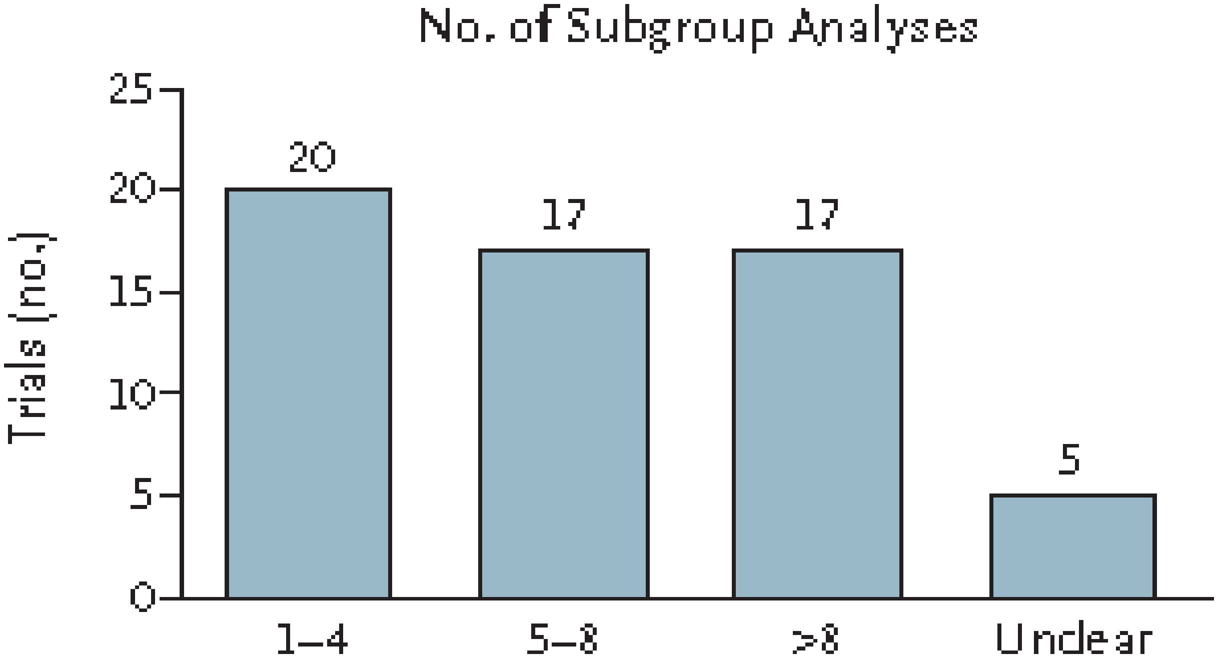
“The essence of tragedy has been described as the destructive collision of two sets of protagonists, both of whom are correct. The statisticians are right in denouncing subgroups that are formed post hoc from exercises in pure data dredging. The clinicians are also right, however, in insisting that a subgroup is respectable and worthwhile when established a priori from pathophysiological principles.”

A R Feinstein, *J Clin Epidemiol* 1998; **51:** 297–99.

Purposes of Post-hoc Analyses (cont.)

- When overall treatment effect is significant in complete study:
  - Identify one or more subgroups where treatment effect more important clinically.
  - Check efficacy benefits in one or more *specific* subgroups where there is *prior reason* to suspect that effect might be reduced or even absent.
- When overall effect is not statistically significant:
  - Identify a subgroup where effect is larger and

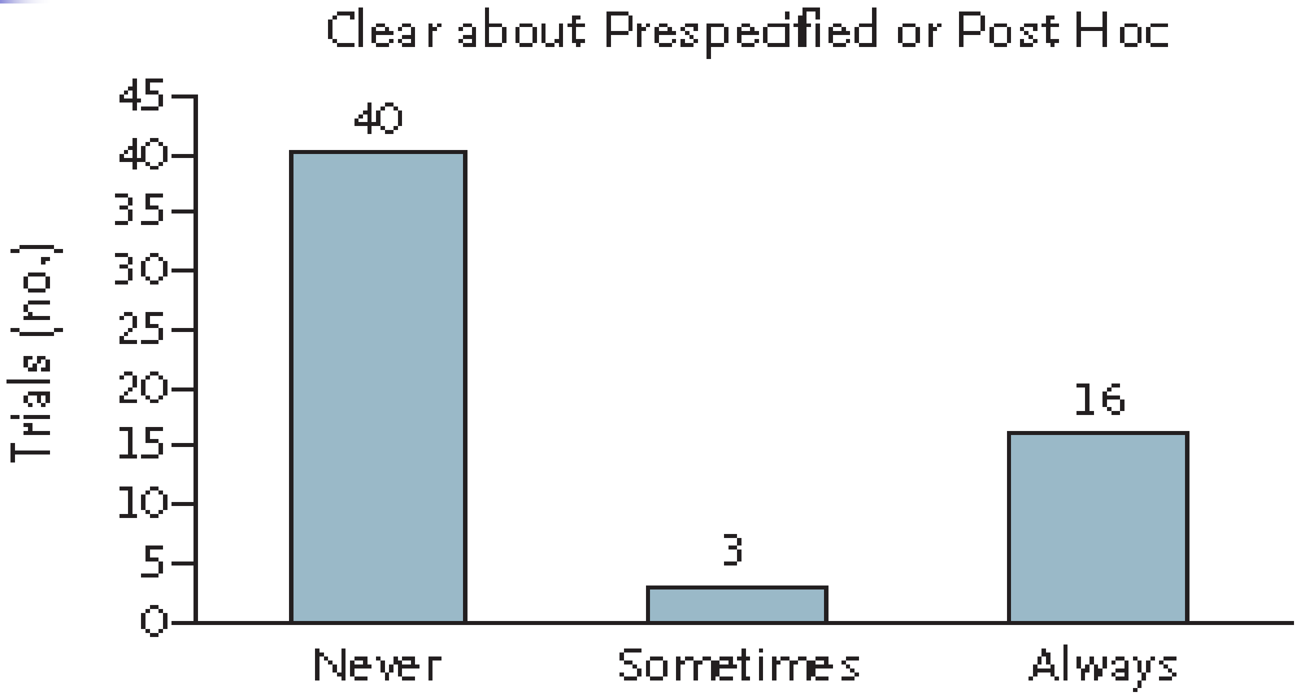
Reporting of Subgroup Analyses from 59 Clinical Trials:



Source: NEJM guidelines (Nov, 2007)

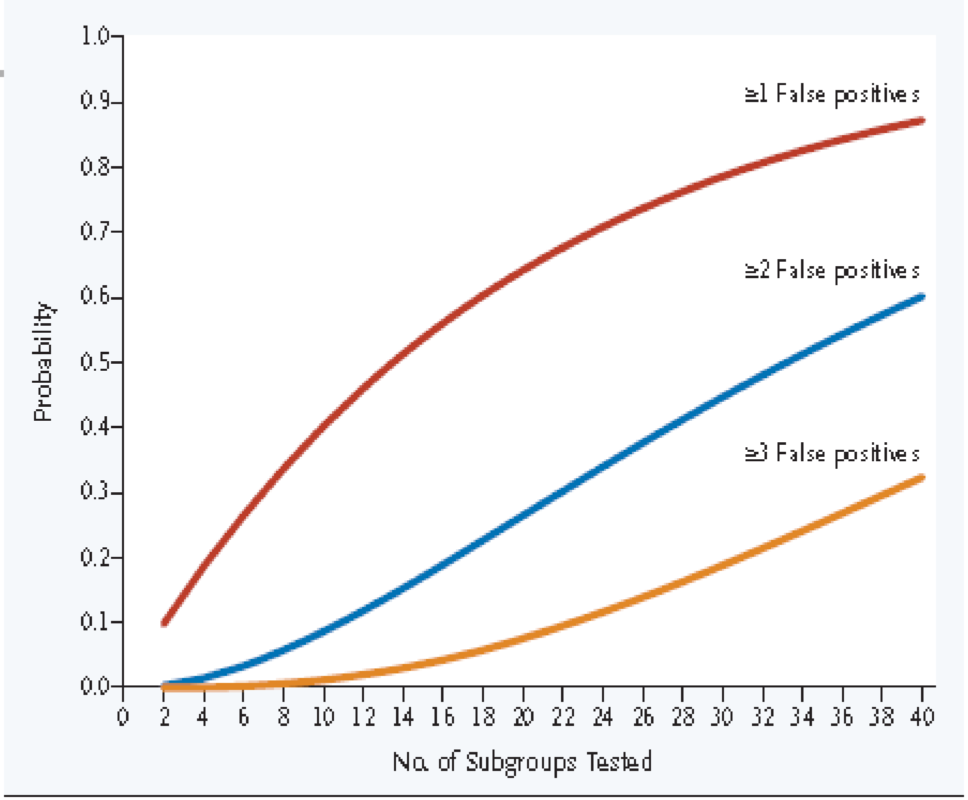
Over 300 subgroup analyses were generated from the 59 Clinical trials!

Reporting of Subgroup Analyses from 59 Clinical Trials:



Pitfalls of Post-hoc Analyses

- Multiplicity
  - With multiple subgroup analyses, probability of a false positive finding substantial.
  - With 10 independent tests (α=0.05), chance of at least one false positive > 40%.



Probability That Multiple Subgroup Analyses Will Yield at Least One (Red), Two (Blue), or Three (Yellow) False Positive Results.

Lagakos (2006) NJM 354;16