Subgroup and Interaction Analyses

C. Michael Gibson, M.S., M.D.

Harvard Medical School
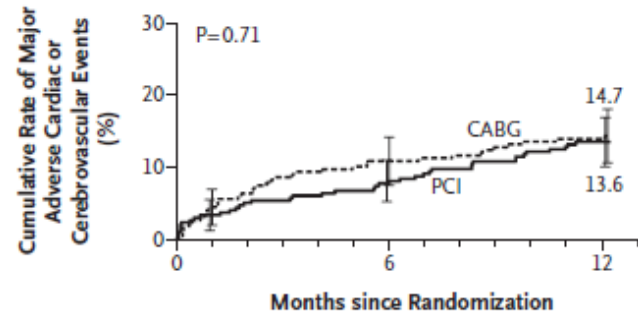
Chairman, PERFUSE Study Group

Founder and Chairman, WikiDoc & WikiPatient, The World's Open Source Textbook of Medicine Viewed 896 Million Times A Year

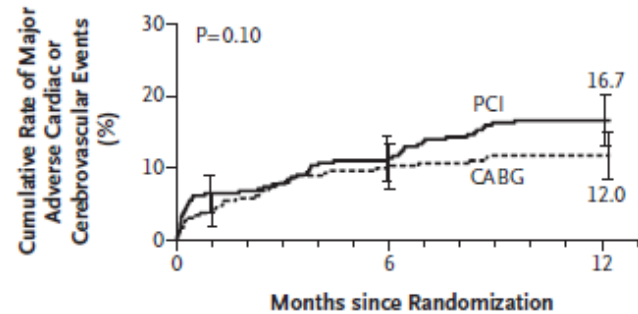# Subgroup and Interaction Analyses

- Almost every clinical trial presents subgroups

    - Within primary manuscript

    - As secondary manuscripts

- What are the pros and cons?

- How do tests for interaction help?

Subgroups often used to inform clinical practice
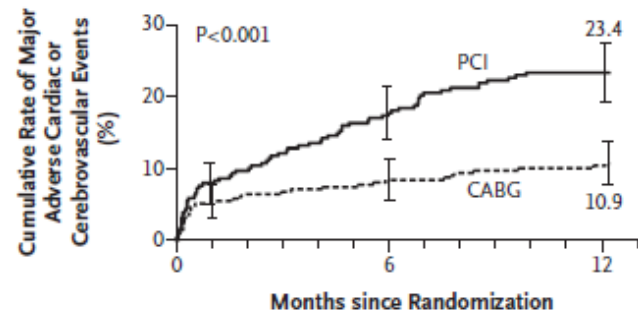
MACCE to 12 Months

CABG (N=897)   TAXUS (N=903)

P=0.0015*

17.8%

12.1%

Cumulative Event Rate (%)

20

10

0

0   6   12

Months Since Allocation

A subgroup may have a completely different result than the primary endpoint

PERFUSE

Event Rate ± 1.5 SE. *Fisher's Exact Test

ITT population

Slide 4
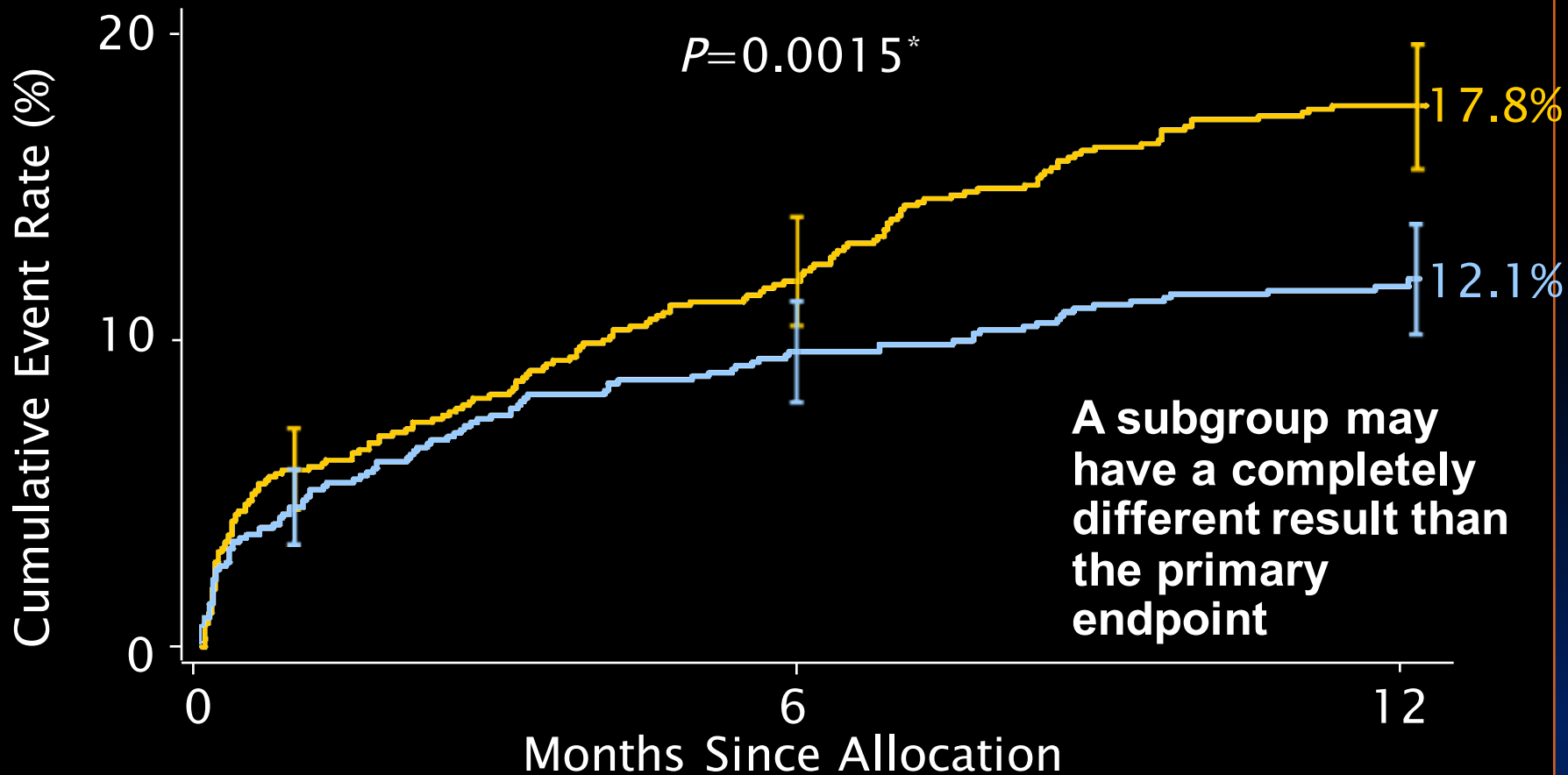
Outcome according to Diabetic Status
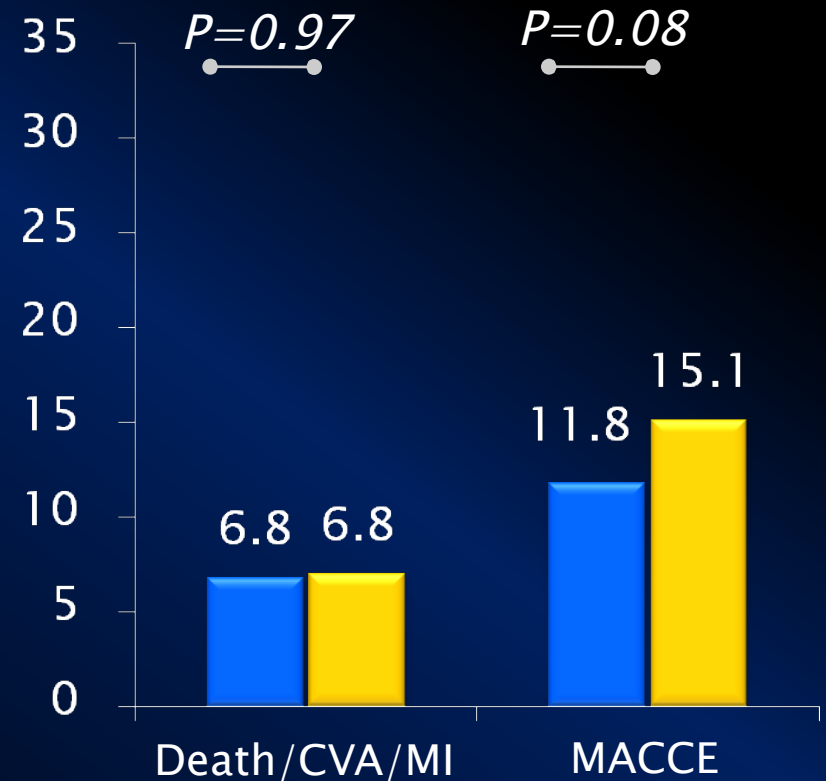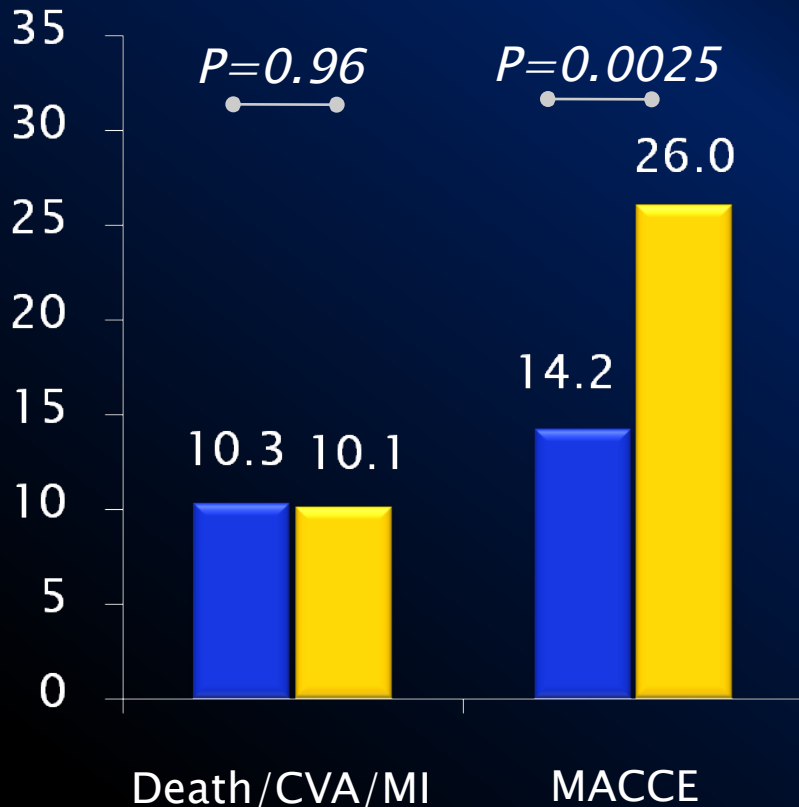
Subgroup and Interaction Analyses

- ## Multiple testing and false positive risk

  - Each time a statistical test is performed there is a chance of false positive (e.g. p value)

  - When multiple related tests are performed, this chance increases according to the number of tests

  - If completely correlated tests – Bonferroni correction estimates the chance of false positive to be

    - P x number of tests

    - If a 0.05 p value is the nominal threshhold, to account for multiple testing divide by number of tests (p/number of tests) for new threshhold

PERFUSE

6 Prespecified Subgroups

Death, MI, Stroke or non-CABG major Bleed

	p-value Interaction
Overall	
Age	
<75	
≥75	0.786
Gender	
Female	
Male	0.356
BMI	
<25	
25-35	
>35	0.637
Radial PCI Volume by Operator	
≤70	
70-142.5	
>142.5	0.536
Radial PCI Volume by Centre	
Lowest Tertile	
Middle Tertile	
Highest Tertile	0.021
Diagnosis at presentation	
NSTE-ACS	
STEMI	0.025

0.25   1.00   4.00

Radial better   Femoral better
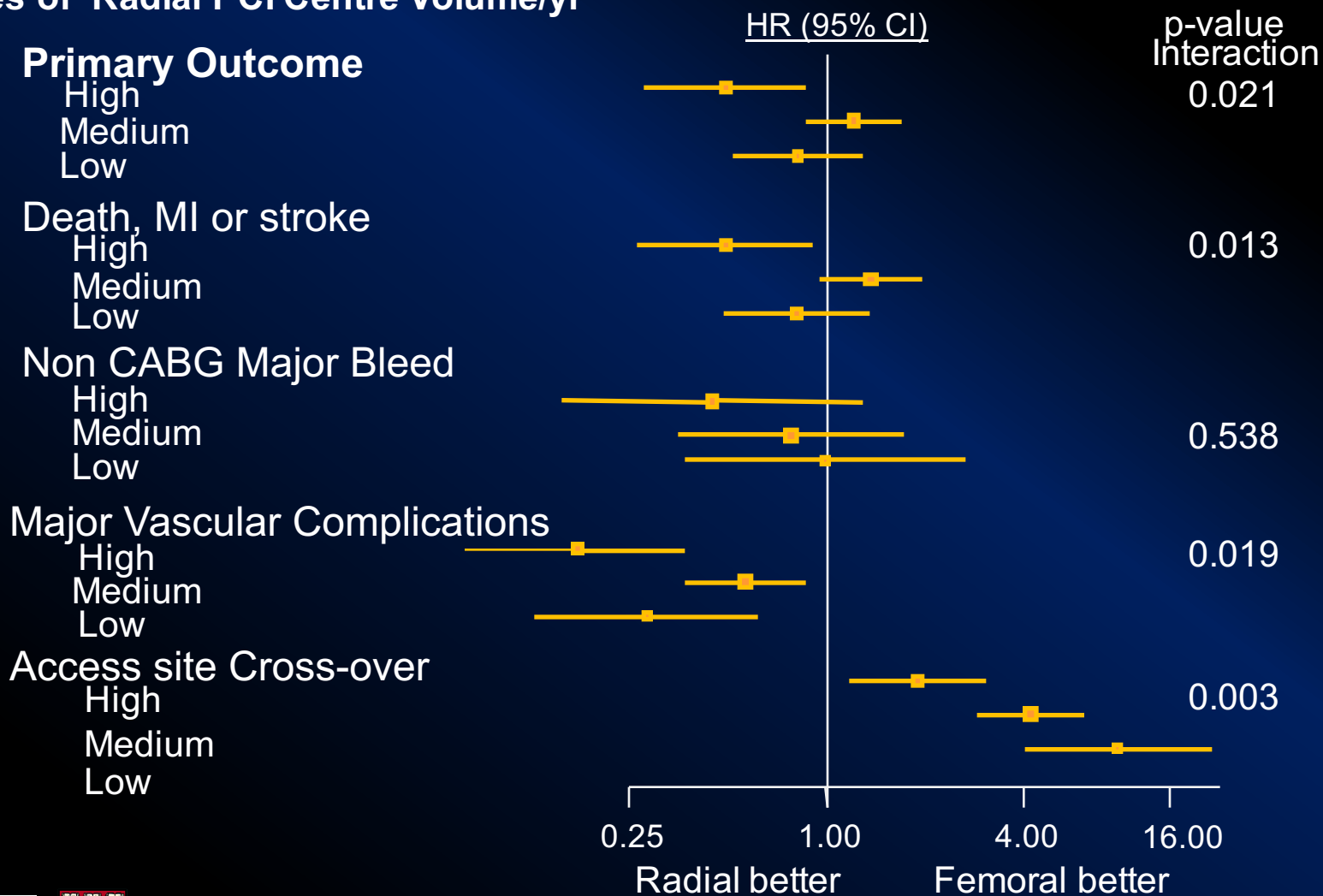
Hazard Ratio (95% CI)

Jolly SS et al. Lancet. 2011 Apr 23;377(9775):1409-20.

PERFUSE

Results stratified by High*, Medium* and Low* Volume Radial Centres

*High (>146 radial PCI/year/ median operator at centre), Medium (61-146), Low (≤60)

Tertiles of Radial PCI Centre Volume/yr

HR (95% CI)

p-value
Interaction

Primary Outcome
High
Medium
Low

0.021

Death, MI or stroke
High
Medium
Low

0.013

Non CABG Major Bleed
High
Medium
Low

0.538

Major Vascular Complications
High
Medium
Low

0.019

Access site Cross-over
High
Medium
Low

0.003

0.25   1.00   4.00   16.00

Radial better   Femoral better

Outcomes stratified by STEMI vs. NSTEACS

	2N	% Radial	% Femoral		Interaction p-value
Primary Outcome					
NSTE/ACS	5063	3.8	3.5		
STEMI	1958	3.1	5.2		0.025
Death, MI or stroke					
NSTE/ACS	5063	3.4	2.7		
STEMI	1958	2.7	4.6		0.011
Death					
NSTE/ACS	5063	1.2	0.8		
STEMI	1958	1.3	3.2		0.001
Non CABG Major Bleed					
NSTE/ACS	5063	0.6	1.0		
STEMI	1958	0.8	0.9		0.56
Major Vascular Complications					
NSTE/ACS	5063	1.4	3.8		
STEMI	1958	1.3	3.5		0.89

0.25 — 1.00 — 4.00

Radial better   Femoral better

Hazard Ratio(95% CI)

Jolly SS et al. Lancet. 2011 Apr 23;377(9775):1409-20.

PERFUSE

Subgroup and Interaction Analyses

- ## RIVAL example

  - 6 prespecified tests of interaction

  - 2 "positive" tests out of 6 at p <0.05 (STEMI, radial volume)

  - If strictly correct, would have required p<0.01 and neither would be "positive"

  - Each was tested according to 5 related outcomes

  - Does this mean no effect if each test is also underpowered?

PERFUSE

Subgroup and Interaction Analyses

- ## How can false positive risk be mitigated

  - Prespecify test of interaction

  - Prespecify subgroups

    - limited number of plausible factors for treatment heterogeneity

    - limited number of endpoints

  - Report the chance of false positive if multiple tests are performed

PERFUSE

Subgroup and Interaction Analyses

- ## Treatment heterogeneity

  - While primary endpoint of a clinical trial examines the mean treatment effect (across a range of patient characteristics)

  - Clinical practice is individualized

  - Interaction terms allow test of whether treatment heterogeneity may be present, according to a single factor

Subgroup and Interaction Analyses

- ## Treatment heterogeneity

  - Treatment effect is not the same in different subgroups

  - Also called "effect modification"

  - Test of interaction (of treatment x subgroup)

  - Represents a challenge in clinical trials

  - Test of interaction reduces multiple testing (compared with test of individual subgroups)

PERFUSE

Subgroup and Interaction Analyses

- ## Interaction types
  - ### No interaction

    OR treatment(diabetes) = OR treatment(no diabetes)

  - ### Quantitative interaction

    Difference in magnitude but not direction of treatment effect

    e.g. OR treatment(diabetes) vs OR treatmen t(no diabetes)

    not equal but same direction, both either >1 or <1

  - ### Qualitative interaction
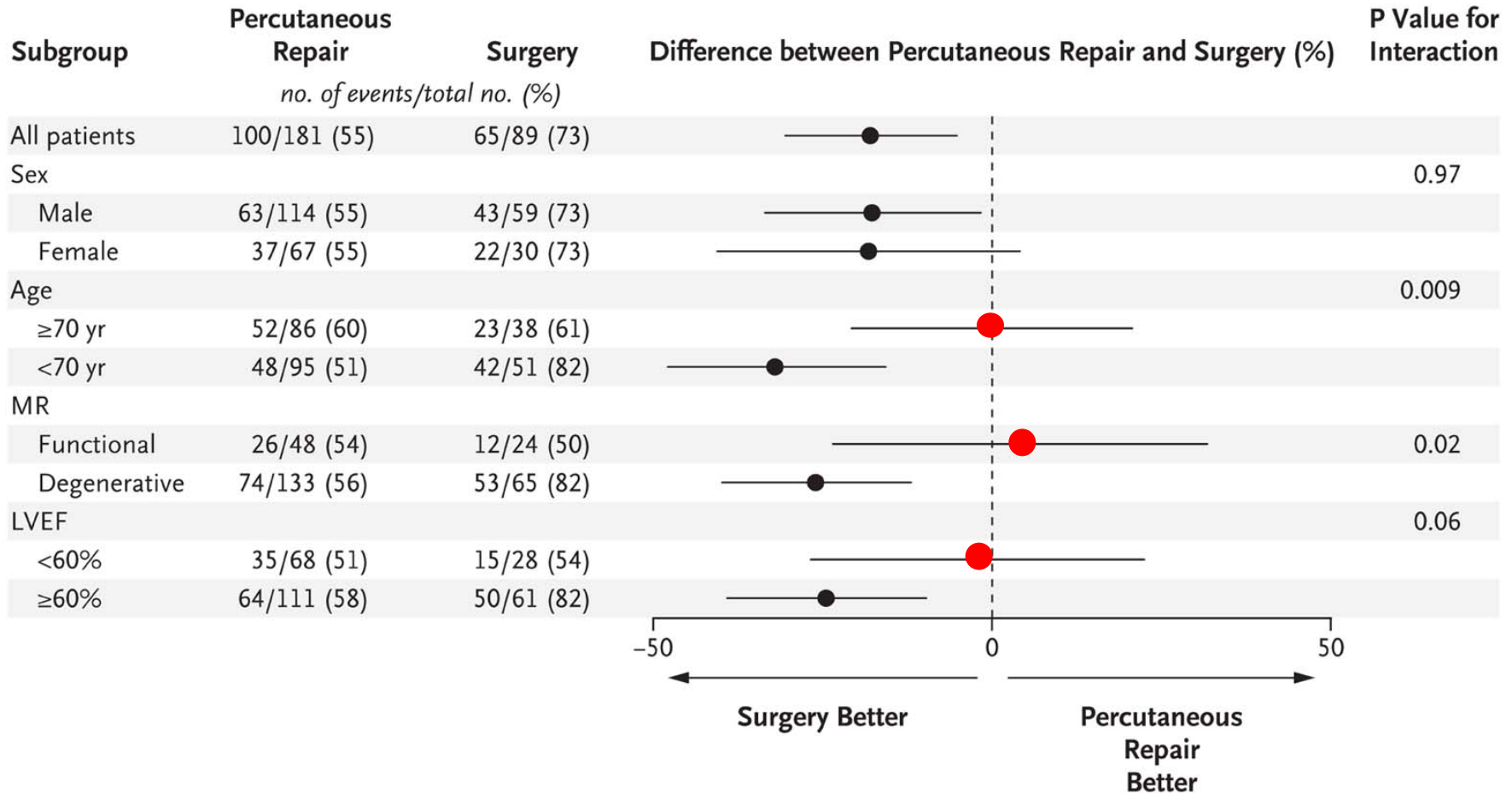
    Difference in direction (benefit in Group A, harm in Group B)

    e.g. OR treatment(diabetes)<1, and OR treatment (no diabetes>1

Subgroup and Interaction Analyses

- 2x2 factorial design a priori specifies and powers for a single test of interaction
- Absent 2x2 design, several *pitfalls* of interaction analysis
  - Multiple testing
  - Power
    - Interaction tests are underpowered
    - Power depends on study sample size and prevalence of risk factor
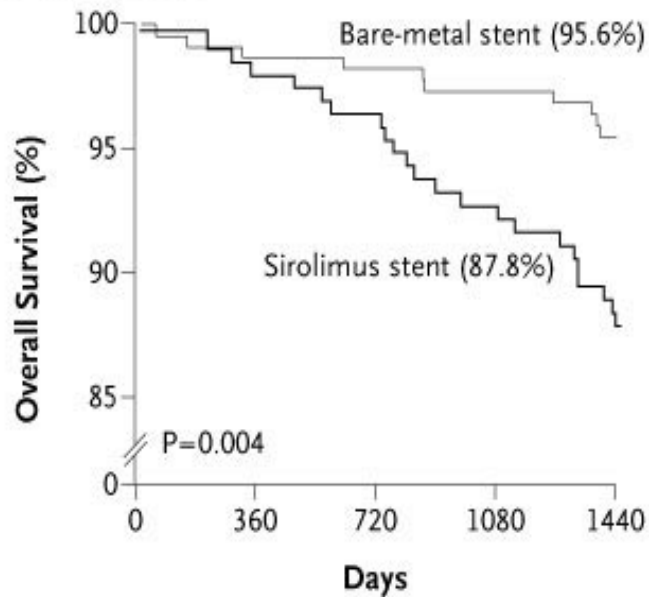 - Hypothesis generating

Subgroup Analyses for the Primary End Point at 12m

Subgroup	Percutaneous Repair	Surgery	Difference between Percutaneous Repair and Surgery (%)	P Value for Interaction
	no. of events/total no. (%)			
All patients	100/181 (55)	65/89 (73)		
Sex				0.97
Male	63/114 (55)	43/59 (73)		
Female	37/67 (55)	22/30 (73)		
Age				0.009
≥70 yr	52/86 (60)	23/38 (61)		
<70 yr	48/95 (51)	42/51 (82)		
MR				0.02
Functional	26/48 (54)	12/24 (50)		
Degenerative	74/133 (56)	53/65 (82)		
LVEF				0.06
<60%	35/68 (51)	15/28 (54)		
≥60%	64/111 (58)	50/61 (82)		

Surgery Better — Percutaneous Repair Better

Feldman T et al. N Engl J Med 2011;364:1395-1406

PERFUSE

Drug-Eluting and Bare Metal Stenting for Diabetes Mellitus

Pooled RAVEL, SIRIUS, E-SIRIUS, C-SIRIUS
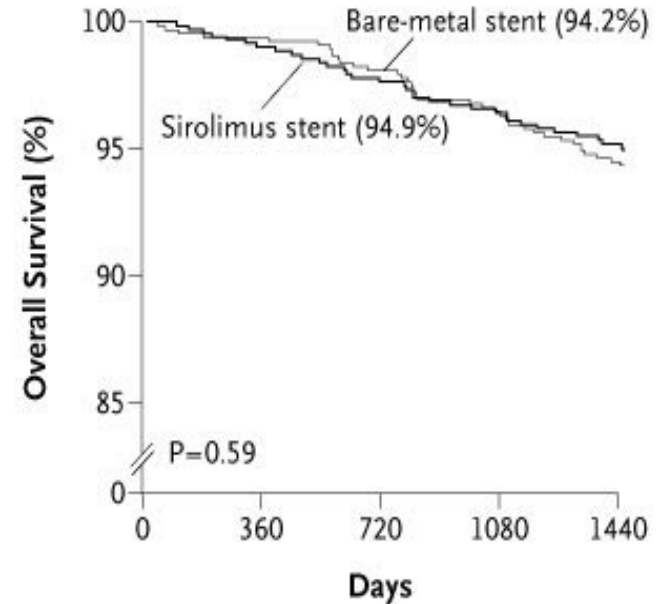
Subgroup and Interaction Analyses

- Subgroup and interaction analyses may be helpful to clinical decision making and areas of future research

- Need careful planning and even more careful interpretation

- It is a good idea to
  - Prespecify a limited number of plausible subgroups
  - Have positive test of interaction precede subgroup analysis
  - Recognize that these are secondary tests
  - Recognize that these tests are usually underpowered, and also subject to false positive risks