Types of Clinical Trials in Cardiology

Rationale for Trial Design: Size, Power, and Subgroup Analysis

Columbia University Medical Center
The Cardiovascular Research Foundation

CARDIOVASCULAR RESEARCH FOUNDATION

COLUMBIA UNIVERSITY MEDICAL CENTER

Basic Clinical Trial Designs

The importance of "Power" in Statistics

- Power Calculations are *critical* when designing studies
  - A RCT is not intrinsically better than an observational study if not adequately powered… in fact, it can be more misleading!
- Power is defined as the ability to be able to statistically detect a difference *when one is truly present*

Basic Clinical Trial Designs

Statistical Power

- **Power is typically set at 80% (or higher)**
  - ***Thus, we accept a 1 in 5 possibility ("fall of the cards") that even if there is an actual difference between 2 stents, we will not be able to find a difference!!!***
  - ***1-power equals the "False Negative" rate***
- **When there is a lot riding on a trial, how much risk can you assume???**
  - **A 10% increase in power (to 90%) will increase sample size!!!**

Basic Clinical Trial Designs

The Importance of Power Calculations

- If I flip a coin twice and it comes up heads once and tails once, does it definitively mean that the coin is fair (or has a 50/50 chance of heads)?

- On the other hand, if I flip a coin twice and it comes up heads twice, does that mean that it will never come up tails (or that it will come up heads twice as often)?

Basic Clinical Trial Designs

The Importance of Power Calculations

- Underpowered studies:
  - When they are negative:
    - Can make two therapies seem similar when in fact differences might exist
  - Confidence intervals can help clarify the picture and determine how certain one can be with the results

Hypothetical Underpowered Trial

- DES A vs. DES B with 500 patients randomized (250 per group)
- 30 day rate of stent thrombosis:
  - 4 events (1.6%) for DES A
  - 4 events (1.6%) for DES B
- Does this mean there are truly no differences between DES A and DES B?

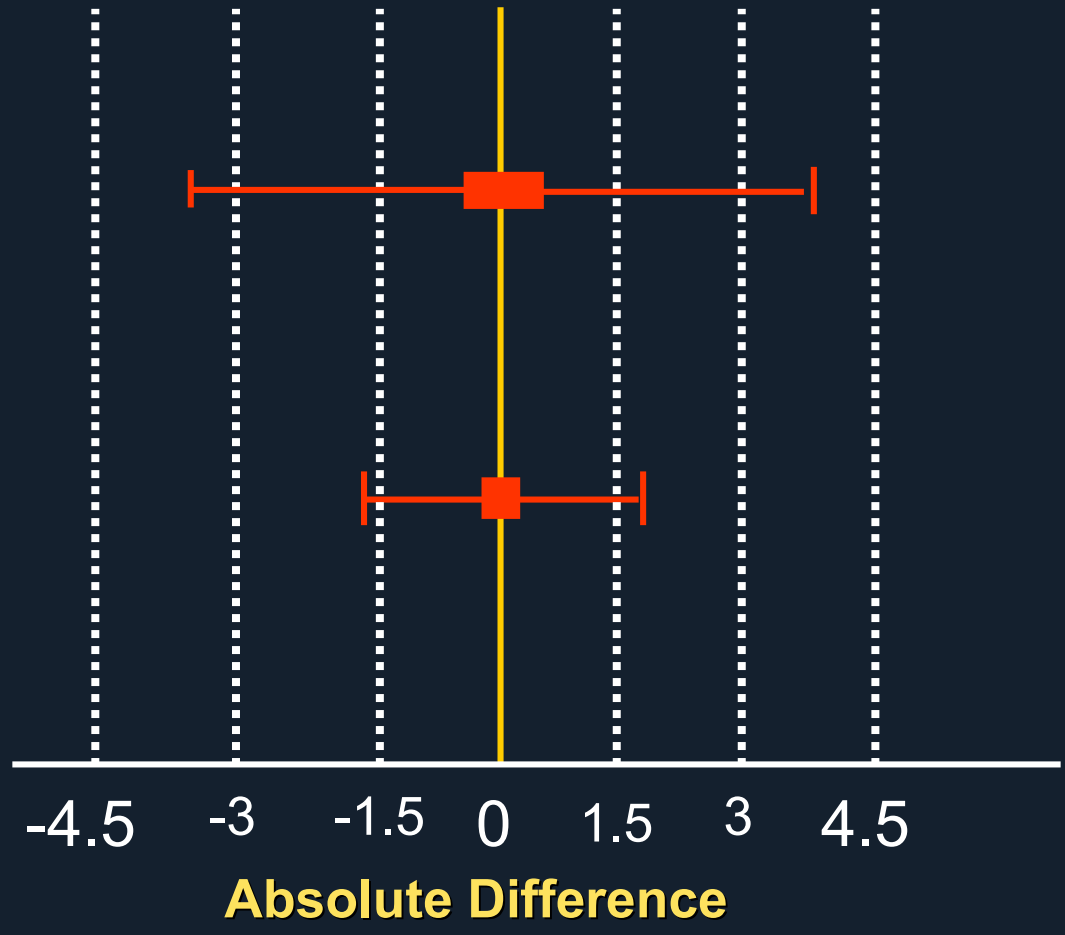
Confidence Intervals of Difference in ST rates between DES A and DES B

In this small trial:

Though there is no difference, we cannot exclude a truly larger difference!

In a larger trial (with 10X as many patients and the same event rates):

There might still be a difference, but we feel more certain of the rates!



Confidence intervals don't lie!

CARDIOVASCULAR RESEARCH FOUNDATION

COLUMBIA UNIVERSITY MEDICAL CENTER

Basic Clinical Trial Designs

The Importance of "Power" in Statistics

- Underpowered studies:
  - When they are "positive":
    - May only get published if results are statistically significant
    - Typically exaggerate treatment effects
    - Even when differences are "statistically significant", the absolute and relative differences are usually overstated

Basic Clinical Trial Designs

Look at Number of Events Too

- Percentages alone can be very misleading – especially when sample size / events are low
- What if the rate of restenosis is 5% with DES A vs. 10% with DES B? This seems like a big difference, but might not be if there were not many overall patients in the study!

Baseline Rate	Total N	Number of Events	95% Confidence Interval
5%	20	1	[0.1%,24.9%]
5%	100	5	[1.6%,11.2%]
5%	1000	50	[3.7%,6.5%]

Basic Clinical Trial Designs

Beware of "Relative Risk"

- For every relative risk increase (or reduction) the baseline risk will determine the absolute risk increase / the number needed to harm

Baseline Rate	Excess Rate (Relative Risk of 2)	Absolute Risk Increase	Number Needed to Harm
0.5%	1%	0.5%	200

Basic Clinical Trial Designs

Relative vs. Absolute Risk

- Even if the relative risk were twice as great, it is important to consider not only relative risk, but *absolute risk* as well
  - If you sell one share of a $1 stock and it then doubles, are you as upset as if you sold one share of Berkshire Hathaway Series A at $134,100 and *it* then doubled?

Basic Clinical Trial Designs

The Importance of Power Calculations

- An extreme example:
- Late stent thrombosis: 0 events with BMS vs. 5 events with DES (p=0.02)
  - The calculated relative risk is *infinite* – is this biologically plausible?
  - Do we think that the rate of late stent thrombosis is truly 0% with BMS?

Basic Clinical Trial Designs

Alpha and the Magical <0.05 Threshold

- Researchers are trained to have an endorphin surge when they see the text "p<0.05"!!!
- But this is somewhat arbitrary… this just means that we accept that there is a 5% or less chance that the results observed (for example showing a difference between two stents) could be due to chance alone
  - Alpha is the "False positive" rate

Basic Clinical Trial Designs

One-sided vs. Two-sided tests

- Statistical convention usually dictates that when we do a trial, we ask up front the question: "Is Stent A *different than* Stent B" rather than "Is Stent A *better than* Stent B"
  - This is a "two-sided" test (admits we are testing the possibility that A could be significantly worse as well)
  - We typically accept a 2.5% false positive error on either side; thus, the total error rate double this value (or $p<0.05$)
- But for non-inferiority tests, we are really testing if A is no worse than B, which *assumes directionality!*
  - This therefore utilizes a "one-sided" test (so we need a more strict threshold of $p<0.025$)

Basic Clinical Trial Design

Sample Size Calculations 101

- What do I need?
  - Baseline assumptions:
    - Event rate in treatment group
    - Event rate in control group
  - Other parameters
    - Balanced or Unbalanced Randomization (1:1 or other)
    - Superiority or Non-Inferiority Hypothesis
    - Power (usually 80% or greater)
    - Alpha (almost always 0.05 for two-sided, 0.025 one-sided)
  - A computer program to crunch the numbers

Basic Clinical Trial Design

Audience Poll

- As overall event rates go down, overall sample size goes…. **UP**
- As the difference between groups increases, sample size goes…. **DOWN**
- As randomization goes from balanced to unbalanced, overall sample size goes…. **UP**
- As power goes up, sample size goes.. **UP**
- As alpha goes down, sample size goes…. **UP**

Surrogate Endpoints in DES Studies

- Why use surrogate endpoints?
  - To reduce sample size and yet be adequately powered
  - To avoid minimize randomizing patients to a therapy which might not provide benefit or possibly cause harm
  - To test new technologies and be able to anticipate their outcomes

Statistical Criteria for Evaluating a Surrogate

1. Strong consistent evidence of treatment differences in each trial
2. Strong relationship with clinical outcome
3. Treatment difference in clinical outcome statistically explained by the surrogate *within* each trial
   [Prentice criterion]
4. Magnitude of treatment difference in clinical outcome clearly linked to magnitude of treatment difference in surrogate *across* trials
   [Hughes criterion]

Additional Key Considerations

- Access to patient level data from multiple well conducted trials is required
- Independent verification by FDA is necessary for FDA to accept results
- Demonstration of acceptable surrogacy for effectiveness does not necessarily imply the same for safety and vice versa

Additional Topics

- Subgroup Analyses
- Landmark Analyses
- Meta-Analysis

Subgroup Analyses

- Patients are not homogeneous and thus, response to treatment may well vary
  - Legitimate to explore in subgroup analyses
- BUT… trials are usually not large enough and lack power to detect subgroup effects
- Many possible subgroups
  - Watch for data dredging / false positives (1 in 20)
- *Do not rely on subgroup P-values; use interaction tests instead*

BARI Trial 10 Year Follow-up

	PTCA	CABG	P-value
All patients (N=1289)	29.0%	26.5%	0.18
Non-diabetics (N=1476)	22.7%	23.0%	Interaction test: p = 0.12
Diabetics (N=353)	54.5%	42.2%	

- Was the trial powered to detect differences within the diabetic subgroup – NO (68% power with this effect size!)
- *Interaction Test Definition:*
  - *Does the overall trial result (which was powered) vary by the diabetic status of the patients?*

NO!!!!

CARDIOVASCULAR RESEARCH FOUNDATION

The BARI Investigators
J Am Coll Cardiol 2007; 49(15):1600-6

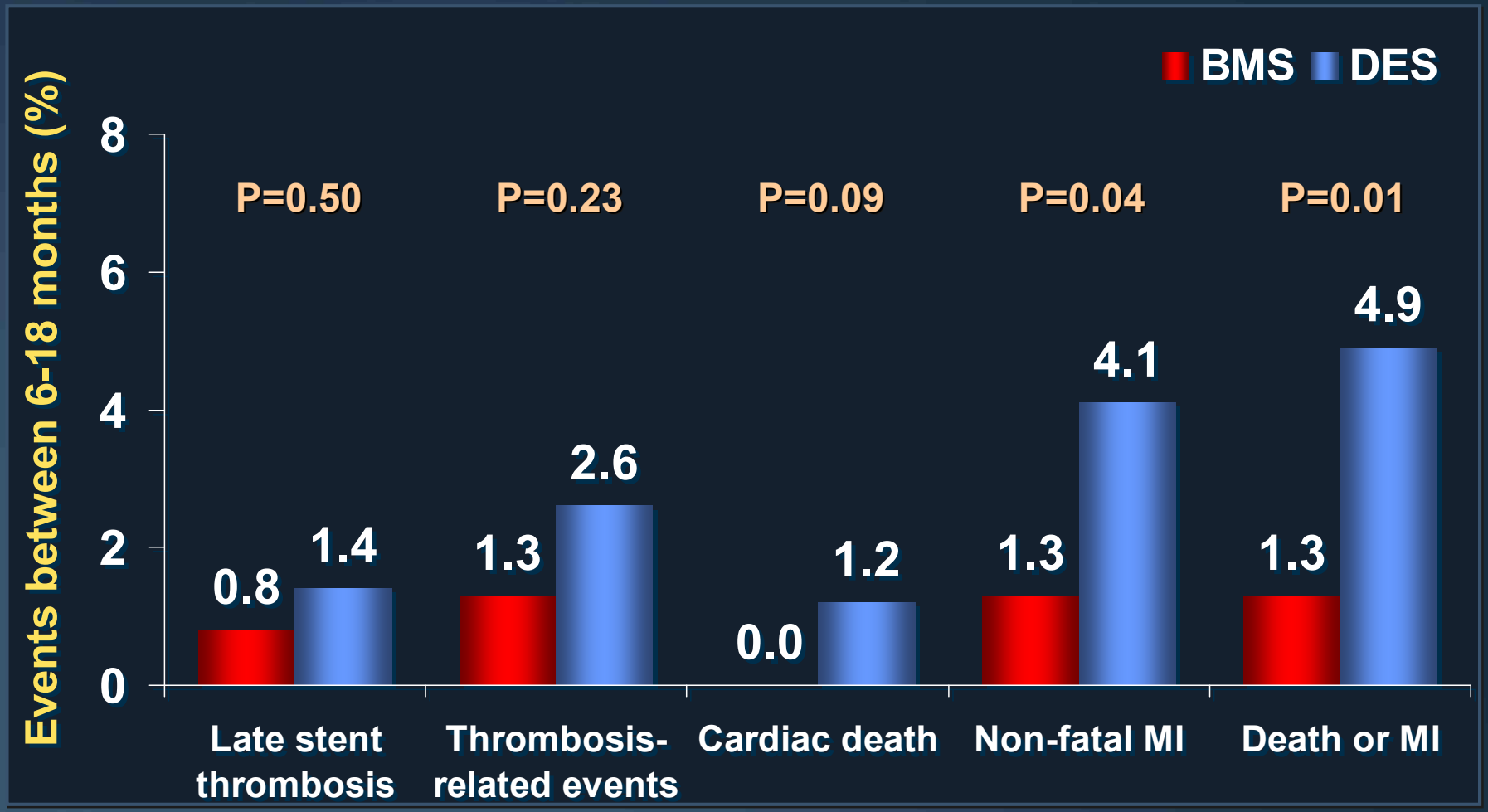
COLUMBIA UNIVERSITY MEDICAL CENTER

Landmark Analyses

- Despite the moniker, they are not (necessarily) "seminal" works
- Most studies assess patients from day of enrollment and assess all outcomes
  - "The landmark" refers to a timepoint from which data is analyzed
  - Earlier events are excluded in these analyses
    - Patients with earlier events are typically not included *unless data on repeated events is gathered/analyzed*
- **Landmark analysis-based information is useful, but selection bias can be a MAJOR issue, as groups are no longer truly "randomized"**

BASKET LATE Trial: 6-18 Mo MACE

N=743 (pts with early events excluded)



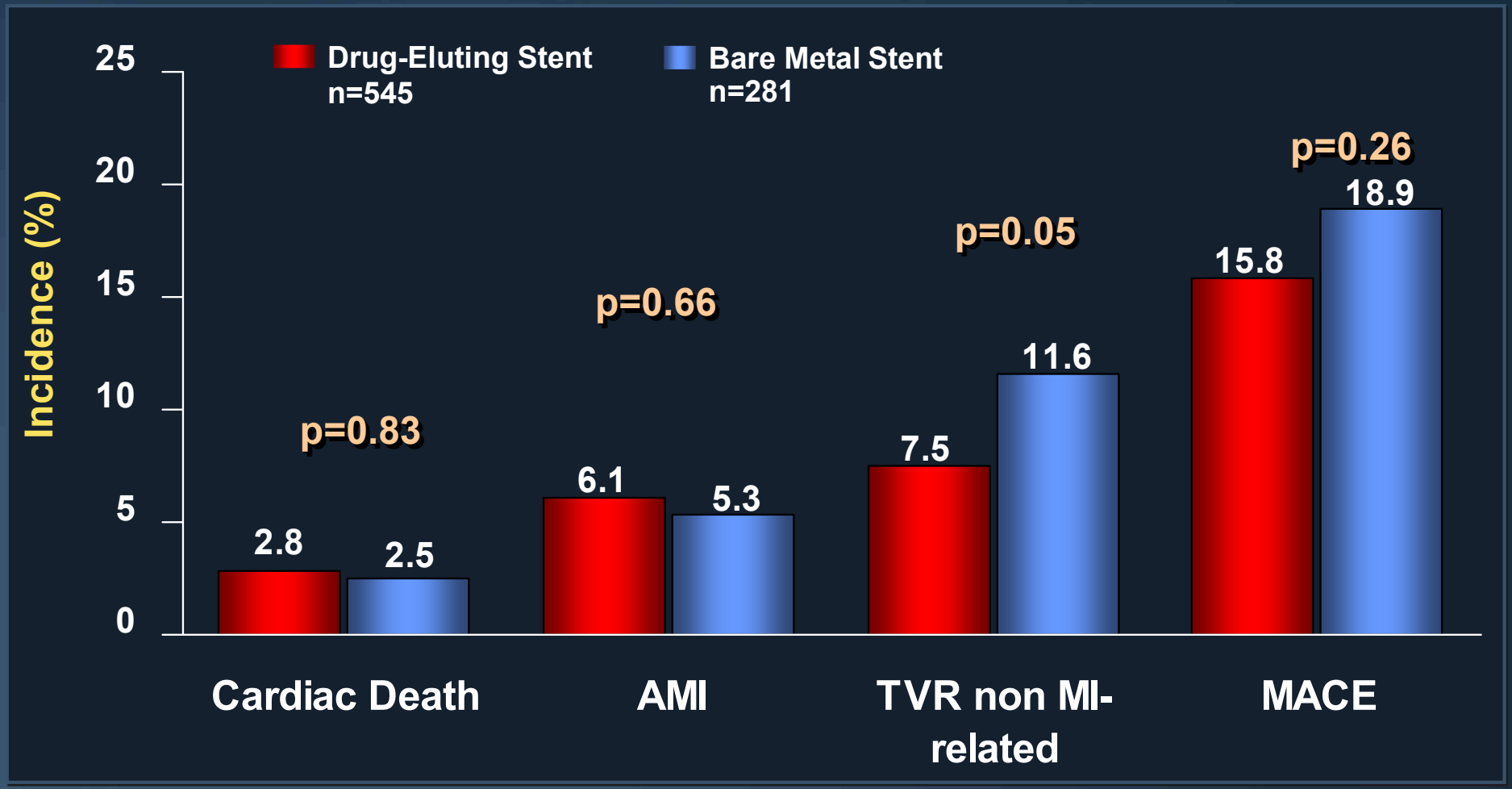
CARDIOVASCULAR RESEARCH FOUNDATION

Pfisterer M. ACC 2006

COLUMBIA UNIVERSITY MEDICAL CENTER

BASKET-LATE: DES vs. BMS

18 Month Clinical Outcomes



Kaiser C. World Congress of Cardiology 2006. Barcelona, Spain.

CARDIOVASCULAR RESEARCH FOUNDATION

COLUMBIA UNIVERSITY MEDICAL CENTER

Meta-Analysis

- A patient asks you: “Doc, what is better, DES or BMS?”
- You either:
  - Don’t know of any data (worst and less common scenario)
  - Know only the last study you read or the study that the last rep told you about (more common scenario)
  - Know all the data inside and out and can put them in a patient-specific context (less common unless your initials are GWS)
- Meta-Analysis aims to synthesize data *quantitatively*
  - But meta-analysis is only as strong as the studies that are included, and still requires a context to be applicable

The Need for Meta-analysis

- Any one study is too small and not generalizable
- Informal literature reviews too subjective

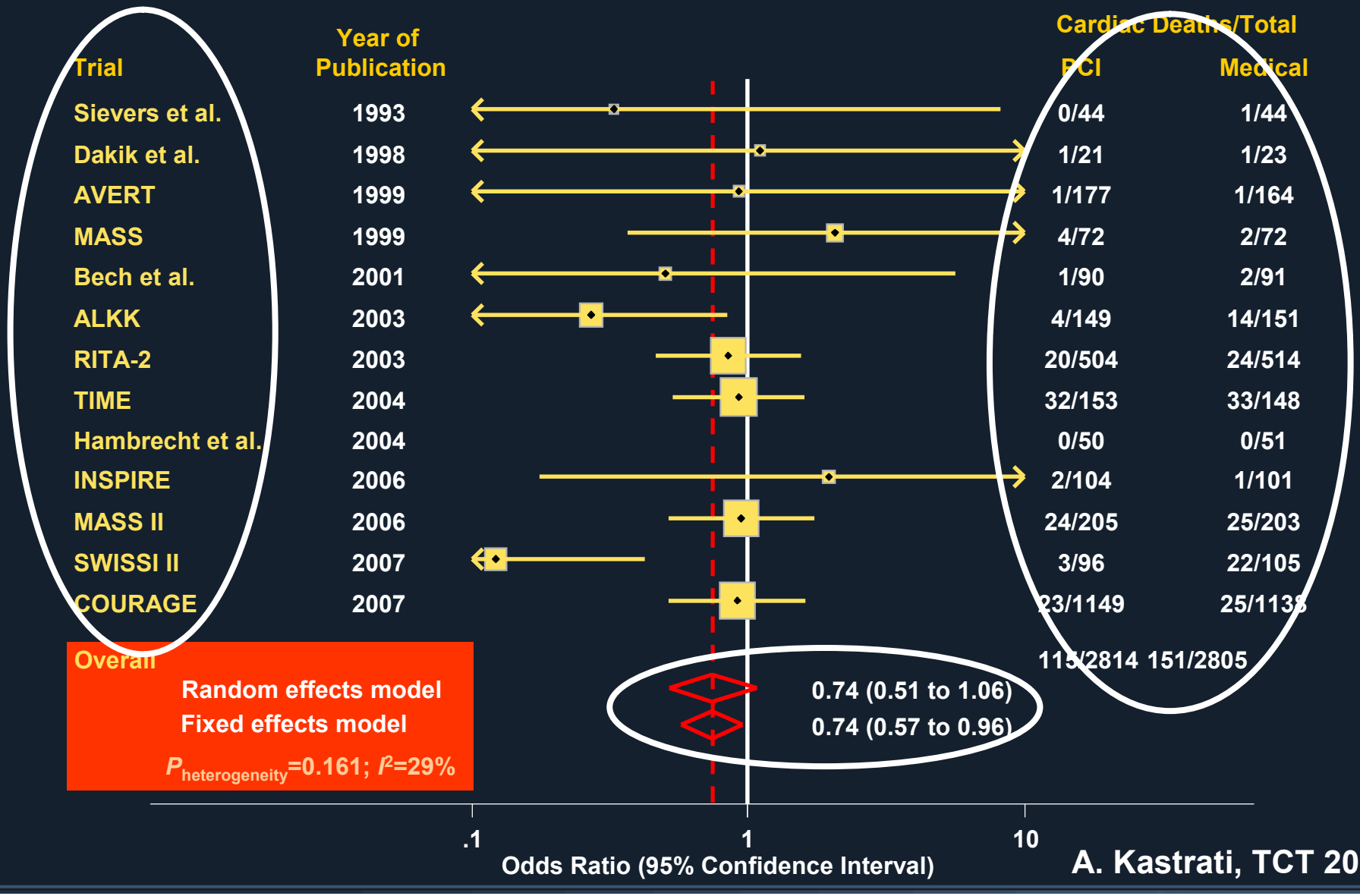
- Combining information from all the trials:
  1. Consistent objective data display
  2. Test of an overall (summary) hypothesis
  3. Estimate an average treatment effect
  4. Investigate if trials are consistent

A Sound Basis for Meta-analysis:

- Similarity of
  - treatments
  - patients
  - endpoints
- All studies included
- Unbiased and comparable study designs in included trials
- Clearly documented inclusion / exclusion criteria and methods
- Sensitivity Analyses

Meta-Analysis: How to Assess

PCI vs. Medical Treatment: Cardiac Death



A. Kastrati, TCT 2007

Conclusions

- Statistics are very powerful tools, but like any tool, they can be misused
- Incomplete understanding and inappropriate uses of statistics can lead to faulty conclusions and mass hysteria (DES thrombosis)
- Always put the data in a clinical perspective
  - The combination of great clinical skills with a knowledge of statistical methodology (and limitations) is a formidable one