



U.S. Food and Drug Administration

Notice: Archived Document

The content in this document is provided on the FDA's website for reference purposes only. It was current when produced, but is no longer maintained and may be outdated.

Data Analysis

Sources of Bias & Error

David L DeMets, PhD

Department of Biostatistics &
Medical Informatics
University of Wisconsin-Madison

Topics For Discussion

- Who & What to Analyze
 - What Patients?
 - What Events?
- Subgroups
 - Baseline Defined
 - Surrogate Defined?
- Missing Data
 - Incomplete Follow-up
- Composite Outcomes
- Non-Inferiority Designs

Patient Withdrawn in Analysis

A. Patient INELIGIBLE

- After randomization, discover some patients did not in fact meet entry criteria
- Concern ineligible patients may dilute treatment effect, so withdraw them
- Withdrawal of ineligible patients, post hoc, may introduce bias

Anturane Reinfarction Trial (1980) *NEJM*

- Randomized, double blind, placebo controlled

	Anturane	Placebo	Total
Randomized	813	816	1629
Ineligible	38	33	71

- Reasons for ineligible
 - 1/3 - time since MI: < 25 days or > 35 days
 - 1/3 - enzymes not elevated
 - 1/3 - other: age, enlarged heart, prolonged hospitalization,
- Number ineligible about the same in each treatment group

BUT

1980 Anturane Mortality Results

	Anturane	Placebo	P-Value
Randomized	74/813 (9.1%)	89/816 (10.9%)	0.20
“Eligible”	64/775 (8.3%)	85/783 (10.9%)	0.07
“Ineligible”	10/38 (26.3%)	4/33 (12.1%)	0.12
P-Values for eligible vs. ineligible	0.0001	0.92	

Reference: Temple & Pledger (1980) *NEJM*, p. 1488

Intention-To-Treat (ITT) Principle

- Anturane example historically important because it established the ITT principle
 - Regulatory (Temple & Pledger, NEJM, 1980)
 - Academia (e.g. May et al, Circulation, 1981)
- ITT Principle
 - Account for all participants randomized
 - Account for all events during follow up
- Modified ITT?
 - Be careful!

B. WITHDRAWAL FOR NON-COMPLIANCE

References: Sackett & Gent (1979) *NEJM*, p. 1410

Coronary Drug Project (1980) *NEJM*, p. 1038

- **Debate: Two Types of Trials**

- 1. Management**

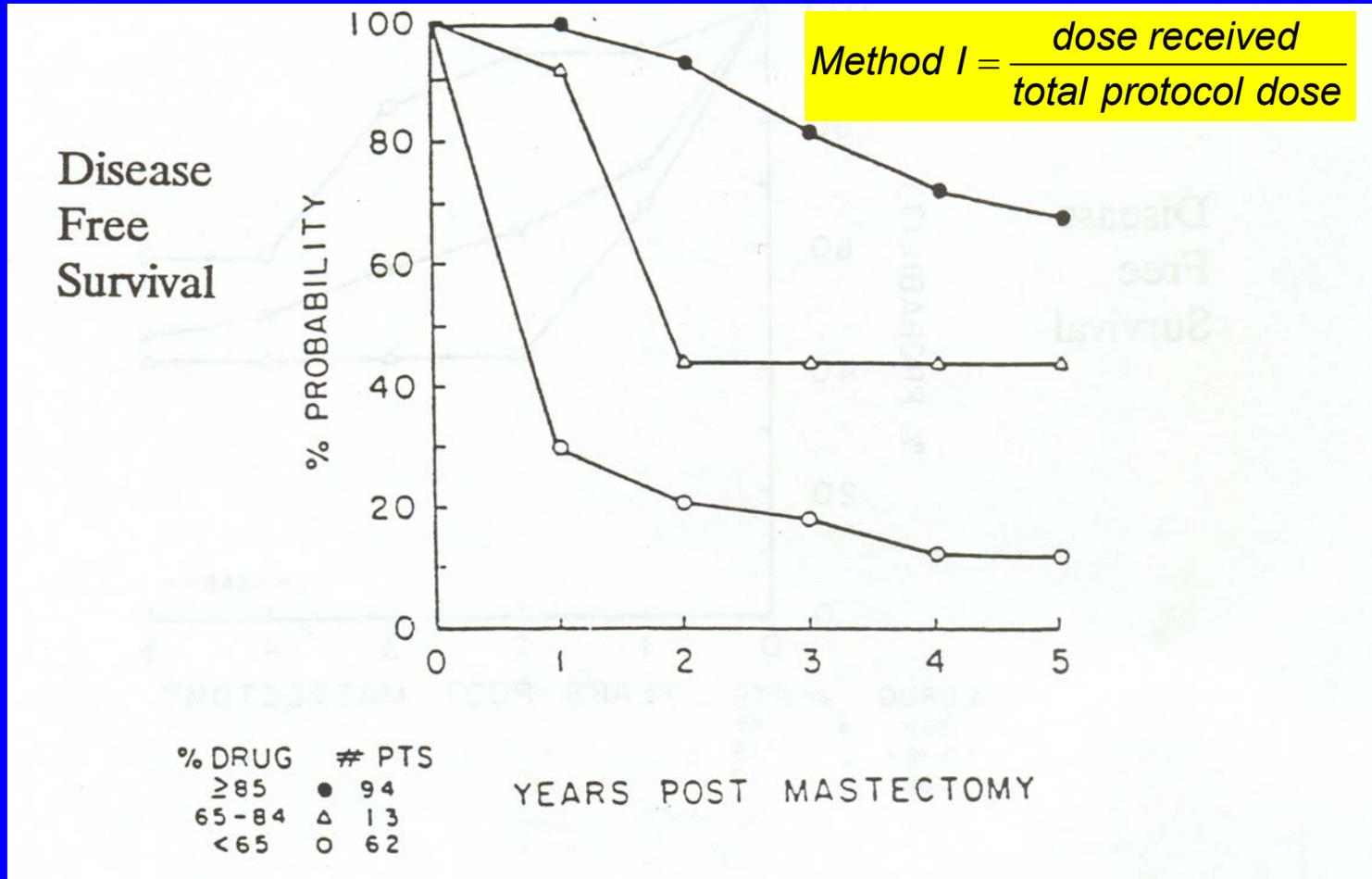
- "Intent to Treat" Principle
- Compare all subjects, regardless of compliance

- 2. Explanatory**

- Estimate optimum effect, understand mechanism
- Analyze subjects who fully comply

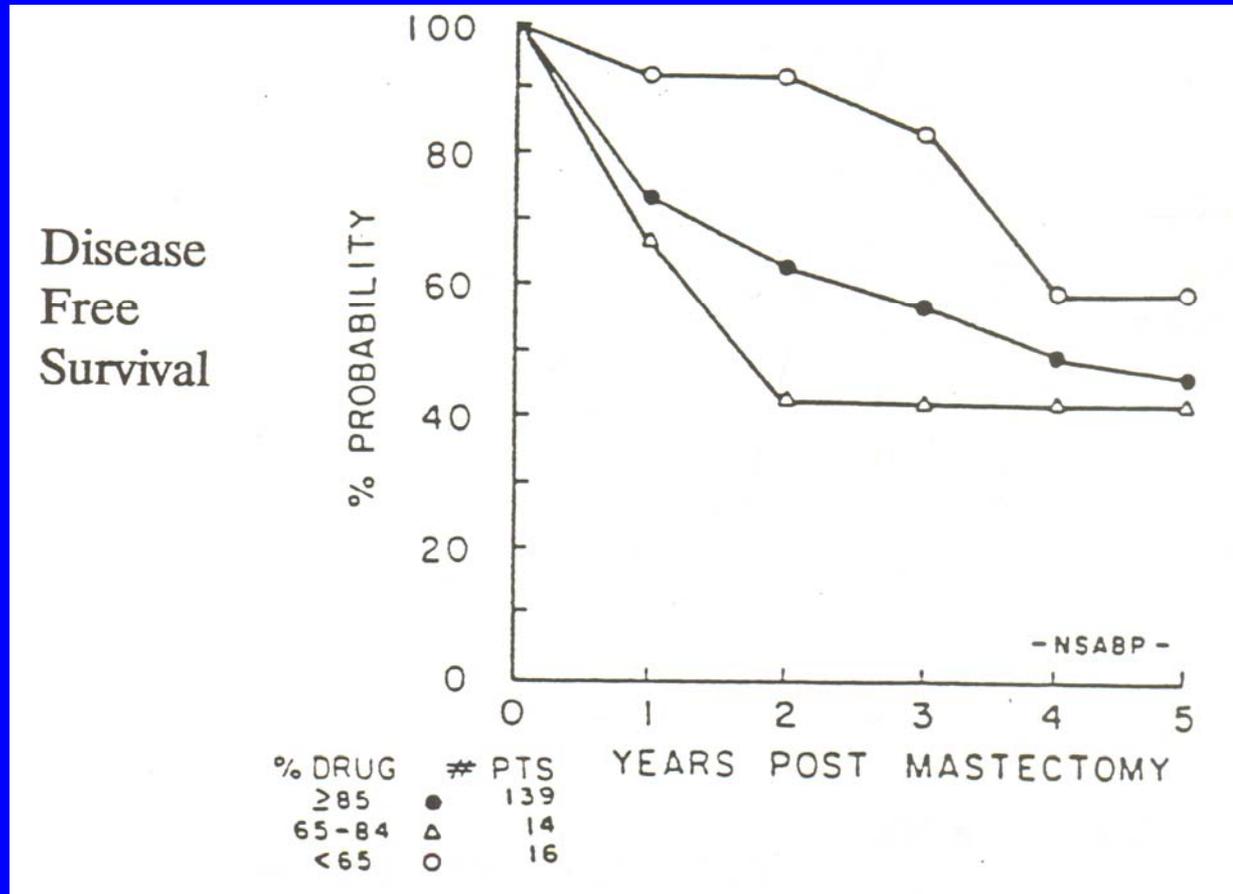
**WITHDRAWALS FOR NON-COMPLIANCE
MAY LEAD TO BIAS!**

Breast Cancer Adjuvant Therapy Probability of Disease Free Survival for Years Post Mastectomy (Method I)



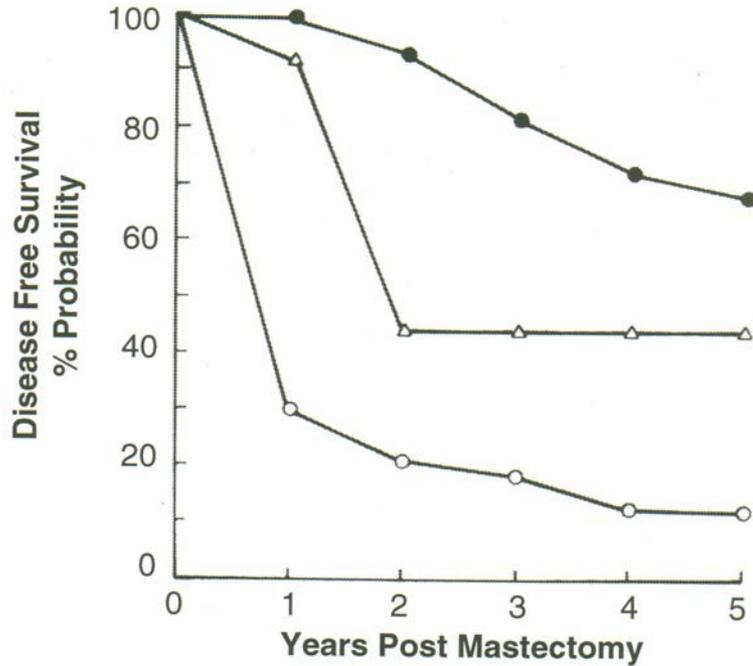
Redmond et al (1983) *Cancer Treatment Report*

Breast Cancer Adjuvant Therapy Probability of Disease Free Survival for Years Post Mastectomy (Method II)



$$\text{Method II} = \frac{\text{dose received}}{\text{dose while on study (possible)}}$$

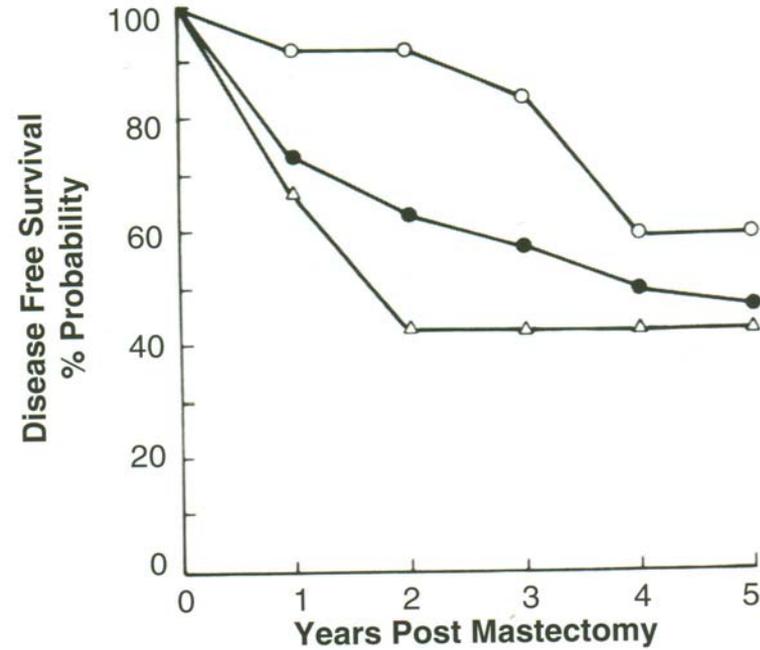
Method I



%Drug	#Pts.
≥85 ●	94
65-84 ▲	13
<65 ○	62

$$\text{Method I} = \frac{\text{dose received}}{\text{total protocol dose}}$$

Method II



%Drug	#Pts.
≥85 ●	139
65-84 ▲	14
<65 ○	16

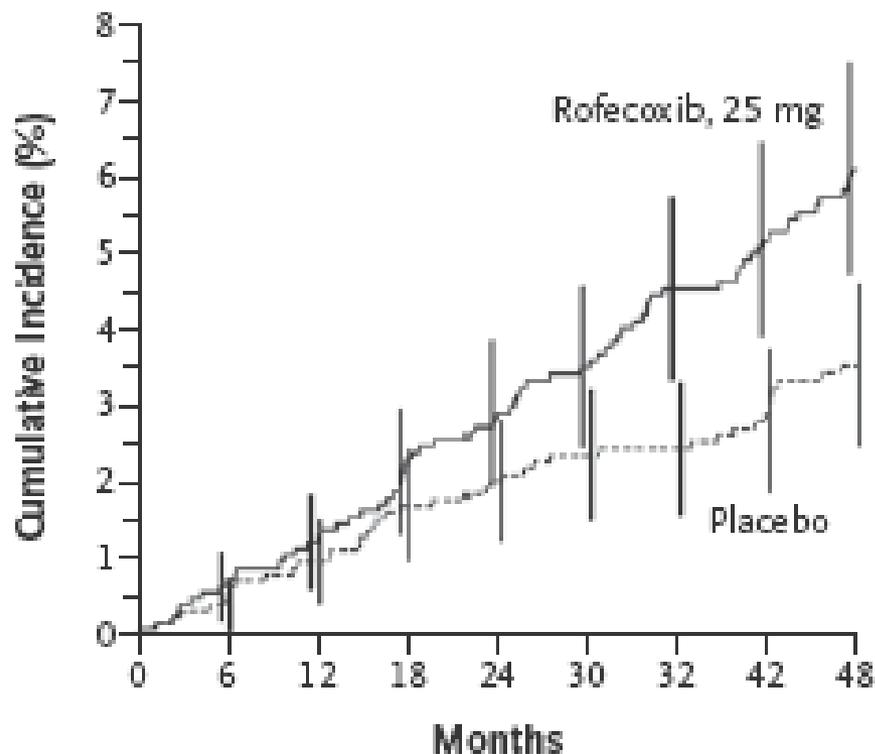
$$\text{Method II} = \frac{\text{dose received}}{\text{dose while on study (possible)}}$$

C: Off Drug \neq Off Study

- ITT requires inclusion of
 - All patients randomized
 - All events during follow up
- Exclusion of either patients or events can lead to bias
 - Direction is not always predictable
- If all events not captured, no way to tell if it makes a difference
- Censoring for going off intervention (e.g. after 7, 14 or 30 days) may be informative

APPROVE Trial

- References
 - NEJM 2005 Primary Paper
 - NEJM 2006 Editorials
 - Lancet 2008 Approve+1
- A trial of Vioxx (Rofecoxib) for colon cancer prevention
- 2005 Paper suggested an increase in CV events
- Debate over 18 month honeymoon



No. at Risk

Rofecoxib	1287	1221	1187	1152	1131	1117	1092	1032	989
Placebo	1300	1247	1224	1189	1173	1157	1133	1071	1027

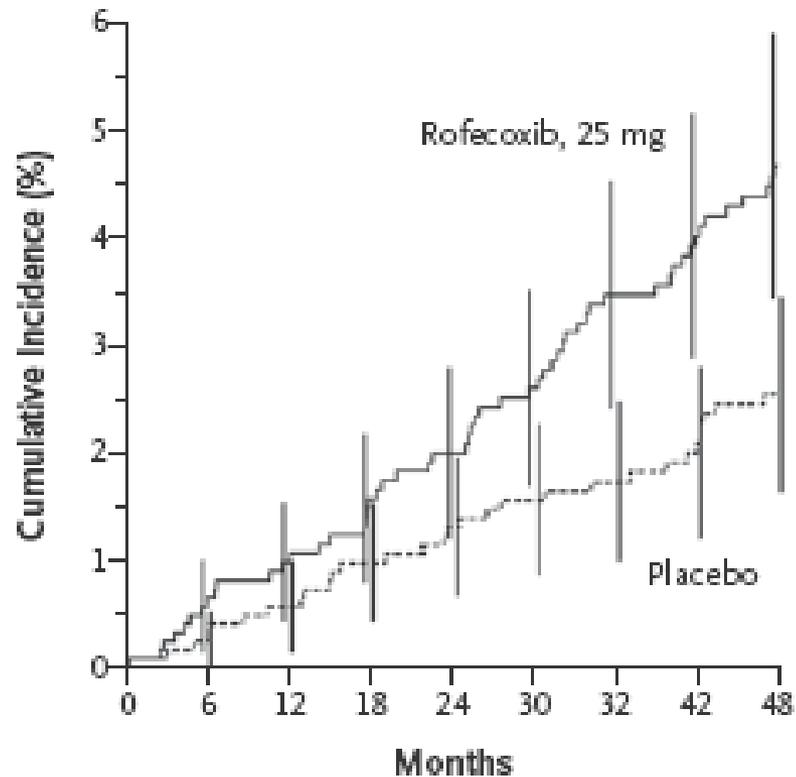
Figure 2. Kaplan–Meier Estimates of the Cumulative Incidence of Confirmed Thrombotic Cardiovascular Events in the Rofecoxib and Placebo Groups, According to the Intention-to-Treat Principle.

I bars represent 95 percent confidence intervals.

Approve + 1

(Lancet, 2008)

- In initial design, patients who went off drug were not followed after 14 days
- Pressures caused sponsor/investigators to conduct an additional year of follow-up on all patients randomized
- An independent analysis was conducted at Univ Wisconsin
- Results with additional year of FU did not confirm the 18 month honeymoon for CV risk



No. at Risk

Rofecoxib	1287	1220	1188	1158	1140	1125	1102	1042	1002
Placebo	1300	1249	1228	1196	1181	1165	1140	1079	1036

Figure 1. Kaplan–Meier Estimates of the Cumulative Incidence of Confirmed APTC Events in the Rofecoxib and Placebo Groups, According to the Intention-to-Treat Principle.

I bars represent 95 percent confidence intervals.

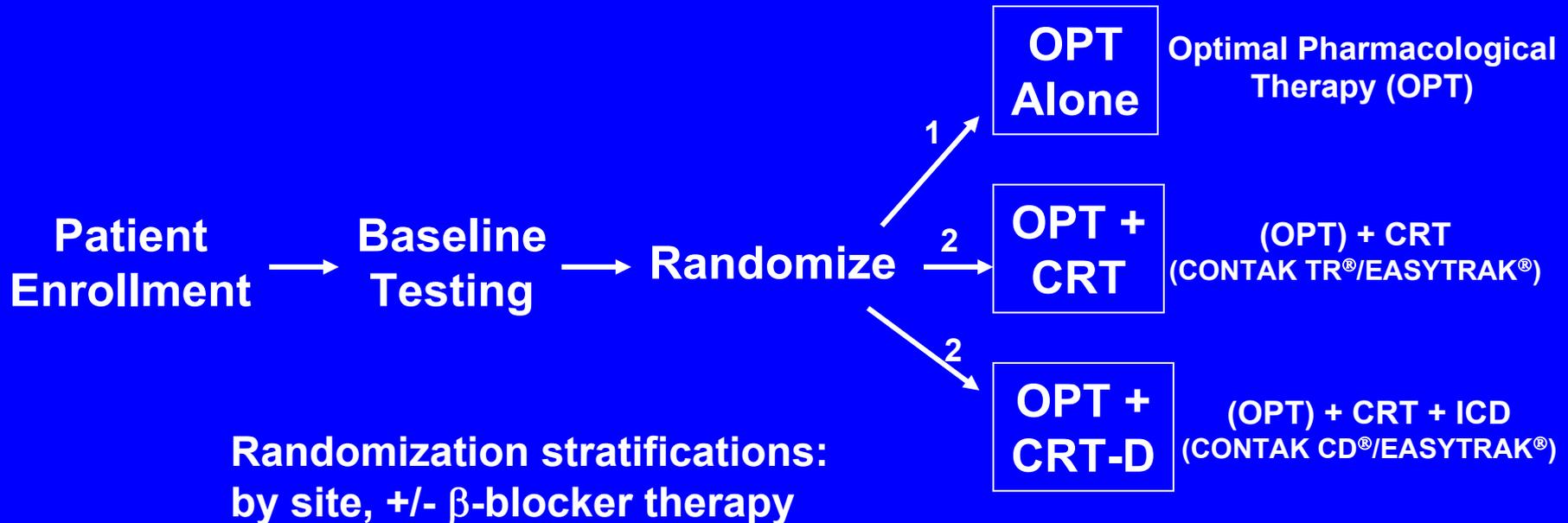
COMPANION

Unexpected follow-up issue

- COMPANION trial was a device trial in CHF patients
- Best medical care vs pacemaker vs pacemaker + defibrillator
- Another device approved during trial
- Patients in best medical care arm withdrew consent; caused follow-up to be censored
- Differential censoring biases analysis

COMPANION (Comparison of Medical Therapy, Pacing, And Defibrillation in Heart Failure): Study Design

Patients randomized 1:2:2
to the following three arms:



Target Time to Implant \leq 2 days from randomization

COMPANION: Endpoints

- Primary Endpoint:
 - Composite of time to first all-cause mortality or all-cause hospitalization analyzed from randomization
 - Hospital emergency or outpatient (unscheduled) administration of IV inotropes or vasoactive drugs for more than 4 hours were considered a hospitalization primary event
 - Later modified to be in hospital over midnight

COMPANION: Data Status

- Trial terminated as planned with follow-up through 12/01/02
- Data indicated a disproportionate withdrawal rate among OPT, CRT and CRT-D (13%, 2%,2%)
- Independent DSMB & blinded Steering Committee recommended:
 - Re-consent withdrawn patients
 - Collect endpoint data and vital status as of 12/01/02
 - Not count elective device admissions as hospitalization

COMPANION: Data Update

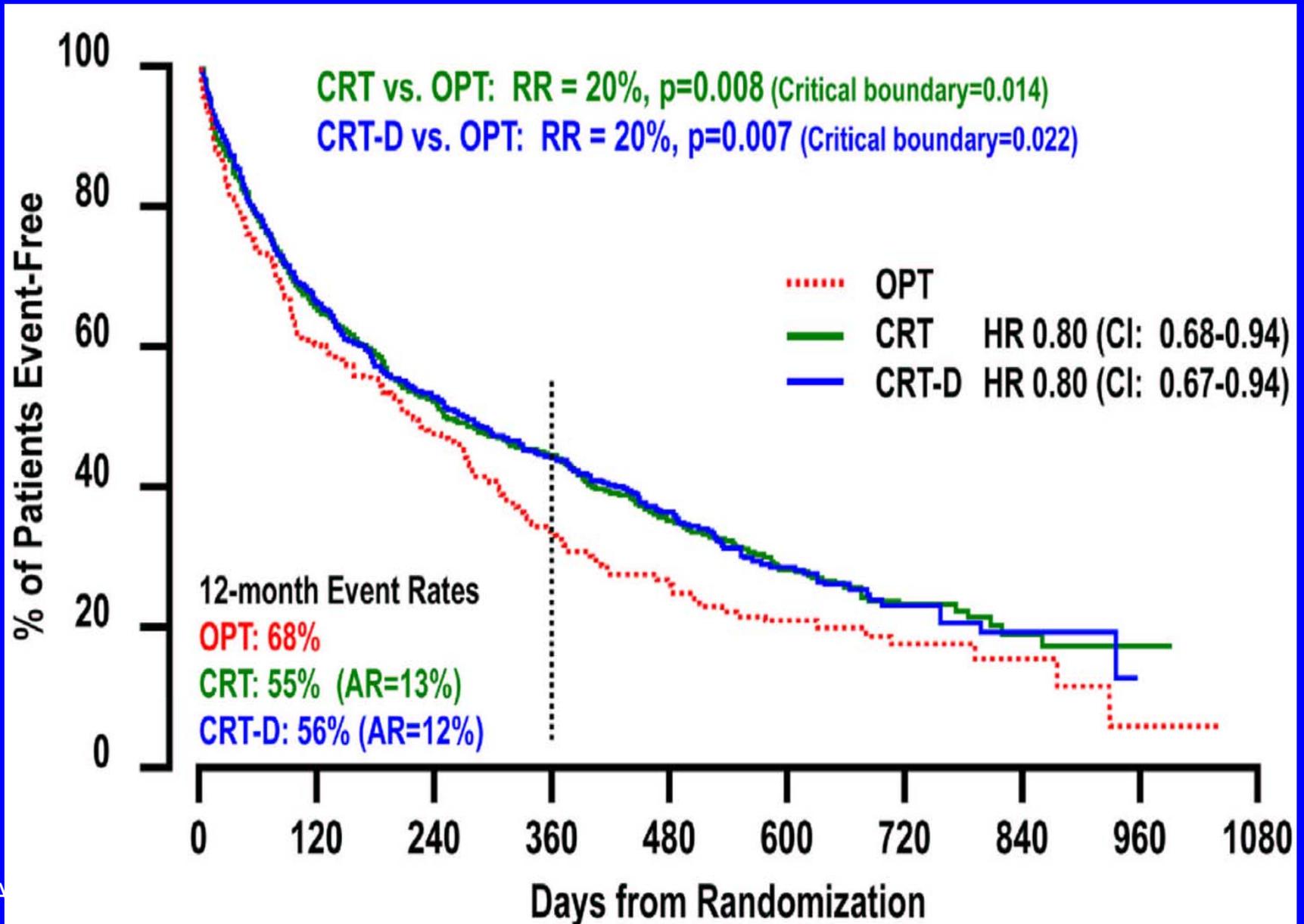
- HFSA Sept 2003 (Final Data)

The process of collecting endpoint data and vital status on patients that withdrew prior to 12/01/02 completed:

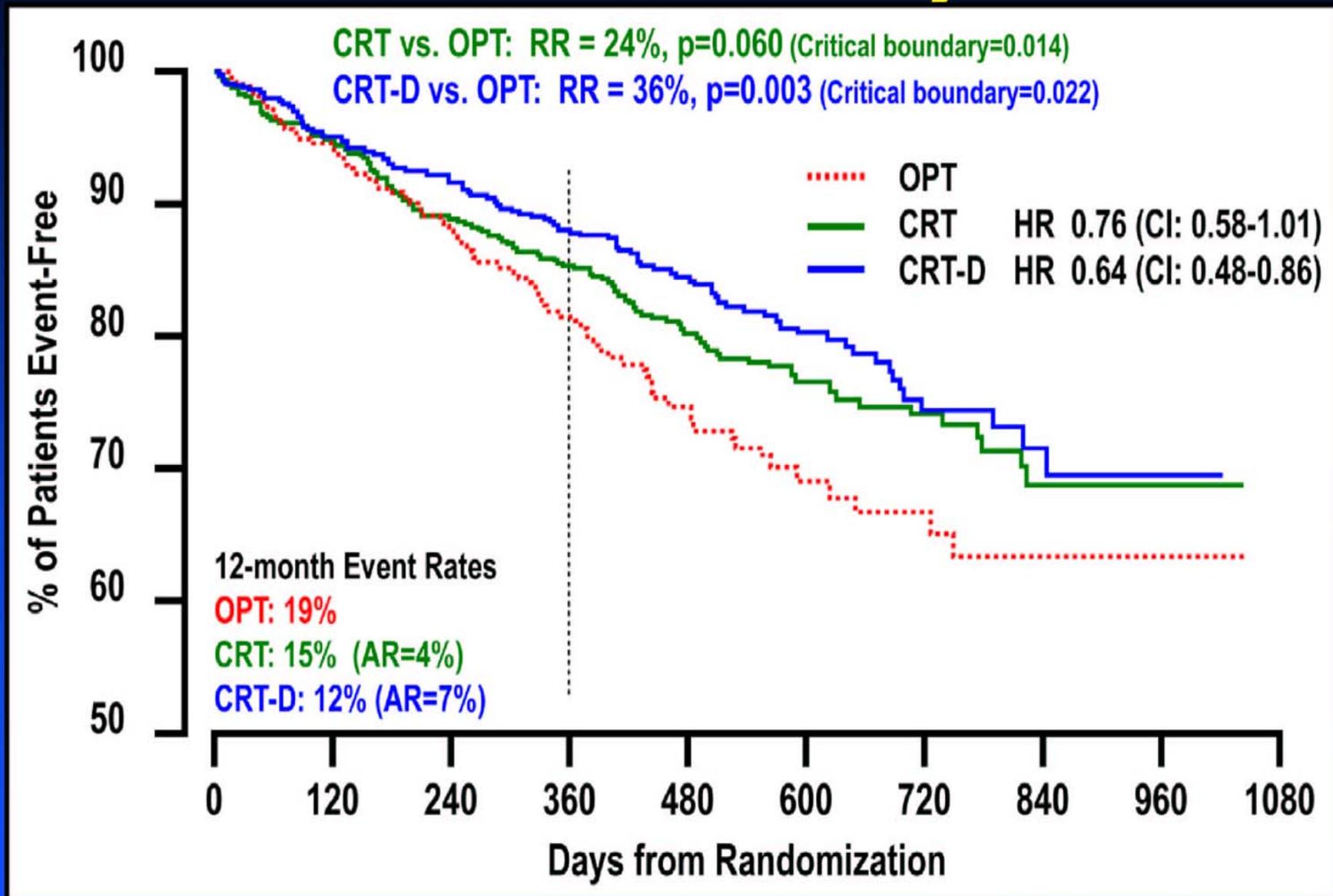
- OPT = 95%, CRT = 99%, and CRT-D = 99%

- Median follow-up times (days) are 442 for OPT, 495 for CRT ($p = .03$), and 479 for CRT-D ($p = .13$)

COMPANION: Primary Endpoint



COMPANION: Secondary Endpoint of All-Cause Mortality



COMPANION

Conclusions & Lessons

- When added to optimal pharmacological therapy in patients with modern-severe LV dysfunction, NYHA class III or IV symptoms and QRS lengthening:
 - CRT or CRT-D reduces mortality + hospitalization
 - CRT-D reduces mortality
- Without additional/completed follow-up, trial would have been difficult to interpret
- Need to plan ahead for consent withdrawal, offering different levels of study withdrawal

Missing Data

- No satisfactory solution
- Need to minimize in design and conduct
- If stuck with missing data, options include
 - Last observation carried forward (LOCF)?
 - Substitution of means
 - Multiple imputation
- Most methods assume missing at random – not likely true

Multiple Imputation

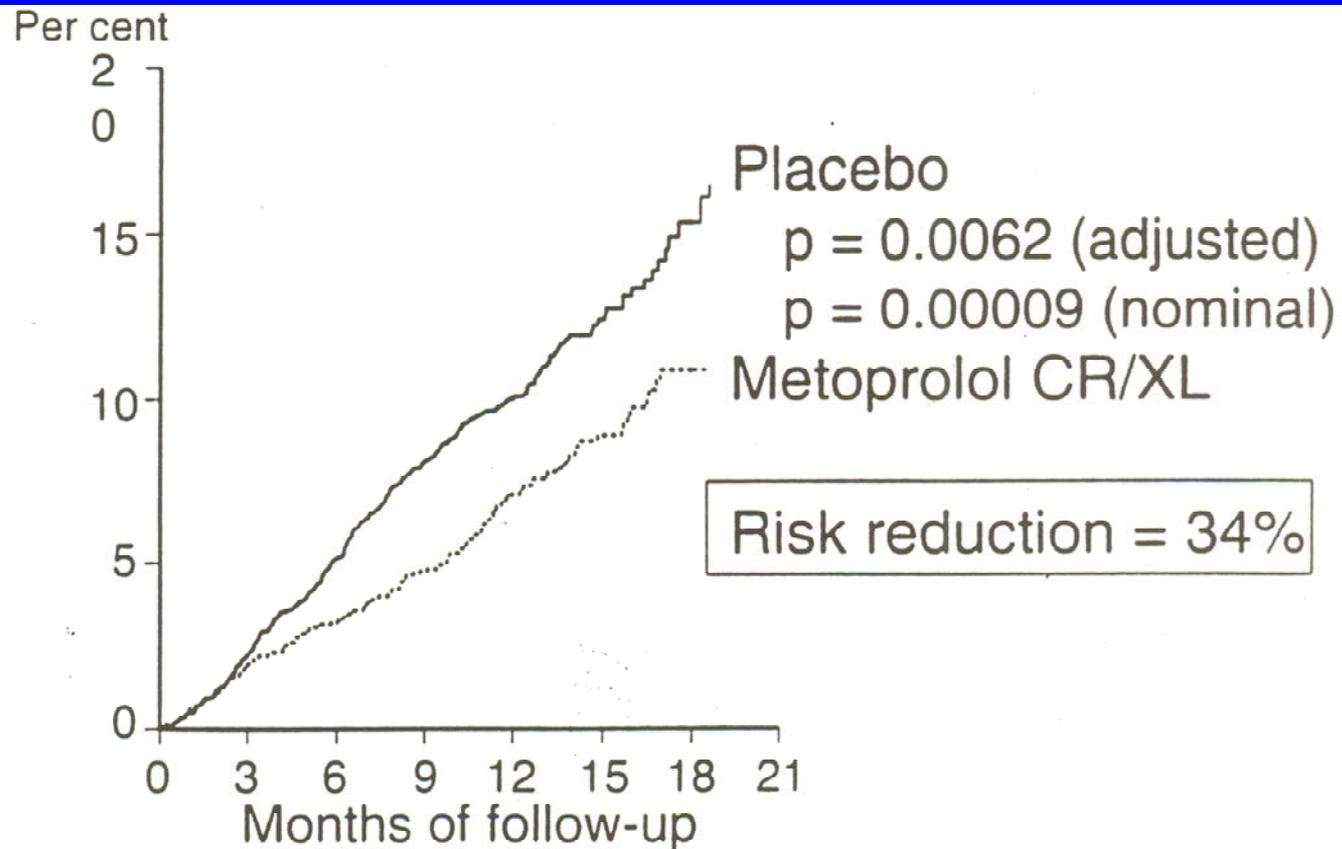
(Rubin, 2006)

- **MI is well established as a valid method of dealing with missing data in the appropriate setting**
- **Each missing data point is replaced by multiple values**
 - **Reflects uncertainty about the correct value**
 - **Allows for standard complete data methods**
- **Standard MI assumes “ignorable” missing-ness**
 - **Ignorable in a particular technical sense**
- **Missing data in a clinical trial probably not missing at random and thus not “ignorable”**

Subgroup Analyses

- Look for qualitative consistency of effect
- Don't expect significance due to smaller sample size
- Focusing on a particular "significant" subgroup can be risky
 - Due to chance, multiple comparisons
 - Results not reliable for small samples
- Results of interest need confirmation

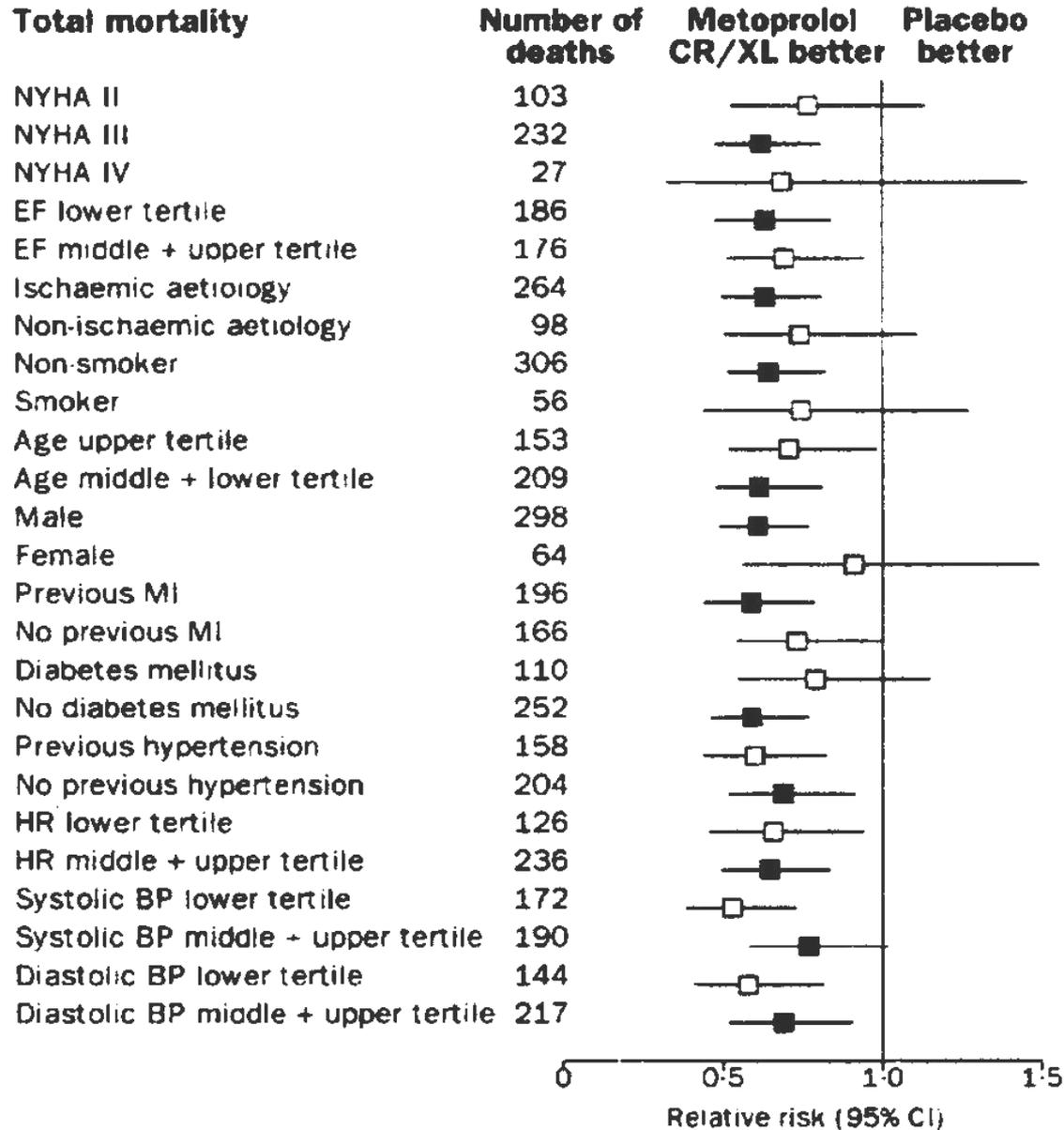
MERIT Total Mortality



Data unblinded by ISaC

The MERIT-HF Study Group. ACC. March 1999

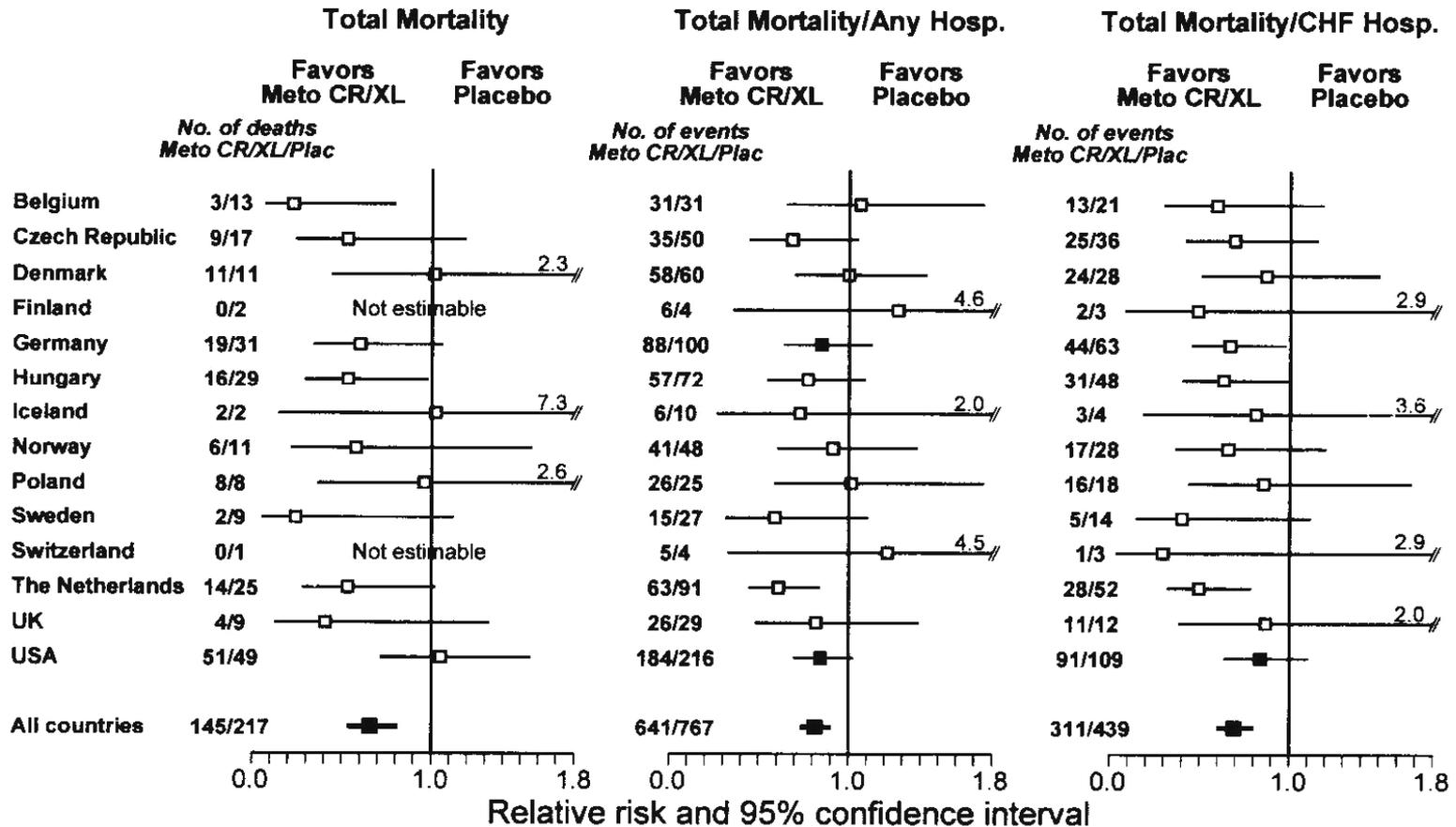
MERIT



MERIT

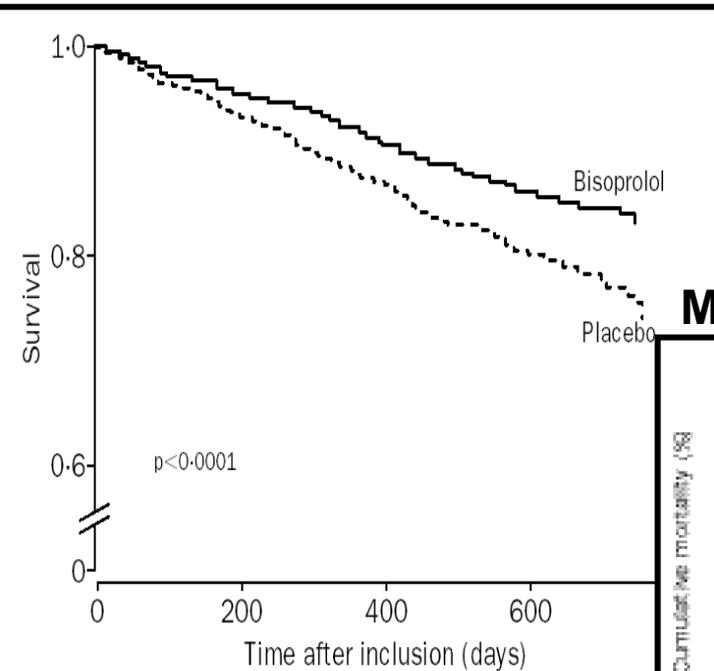
(AHJ, 2001)

All Patients Randomized

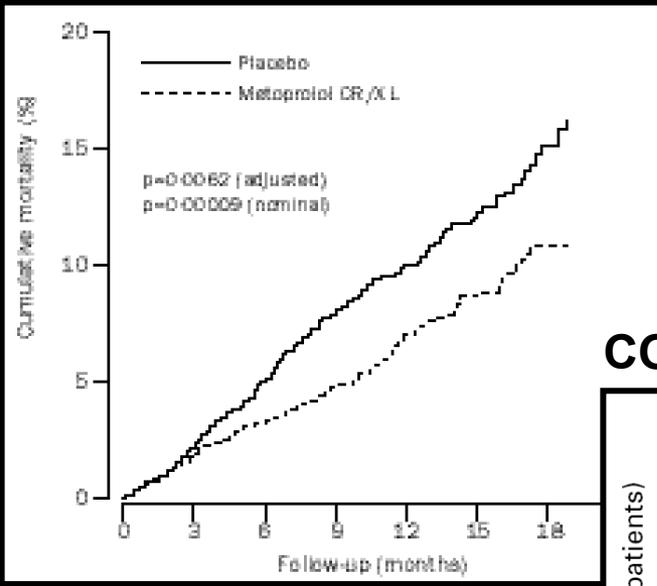


BETA-BLOCKER HF TRIALS

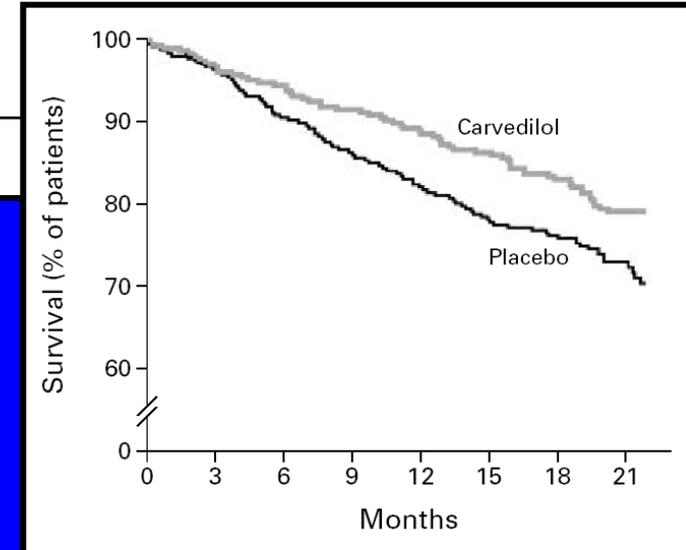
CIBIS-2 1998



MERIT-HF 1999



COPERNICUS 2001

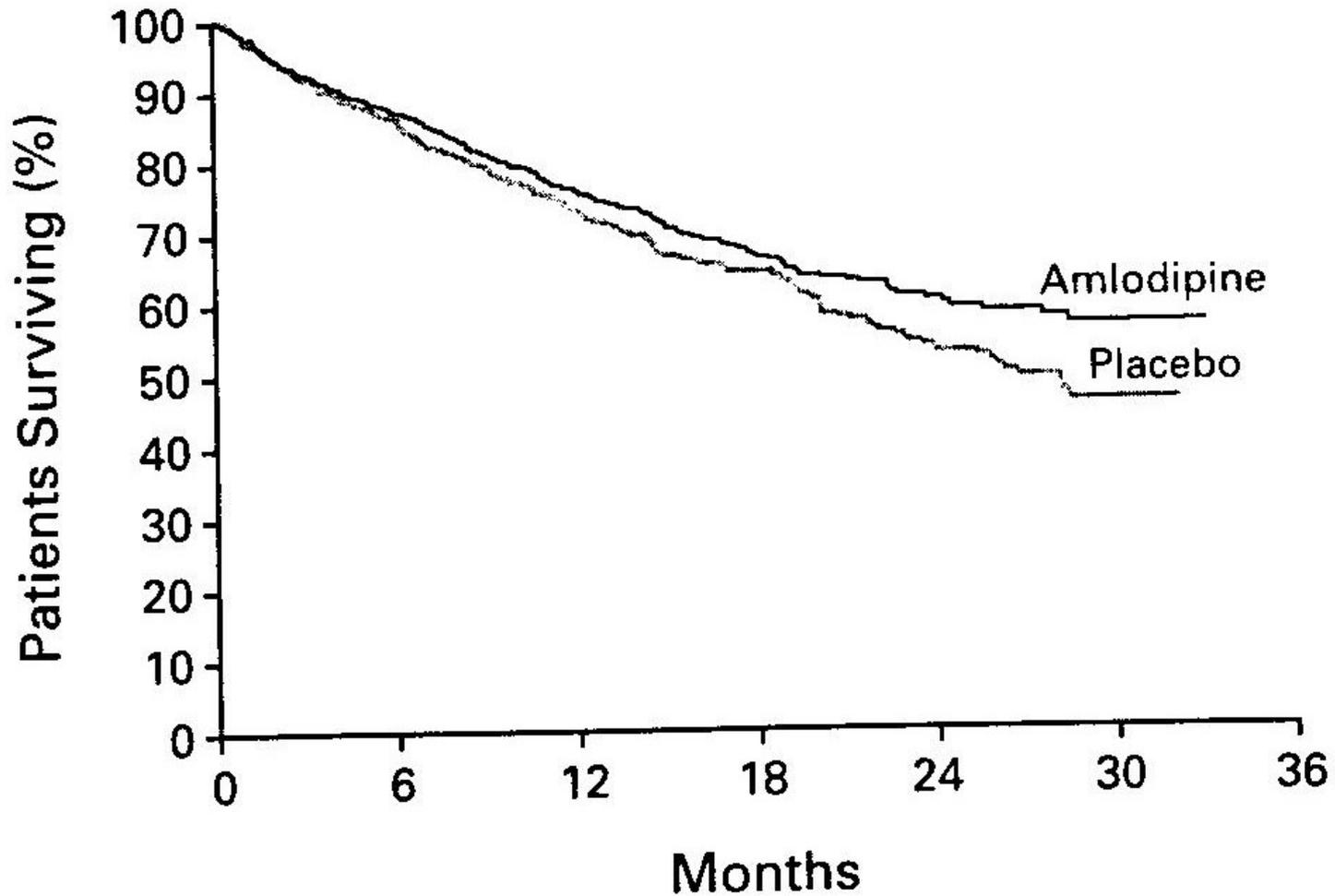


Praise I

Ref: NEJM, 1996

- Amlodipine vs. placebo
- NYHA class II-III
- Randomized double-blind
- Mortality/hospitalization outcomes
- Stratified by etiology (ischemic/non-ischemic)
- 1153 patients

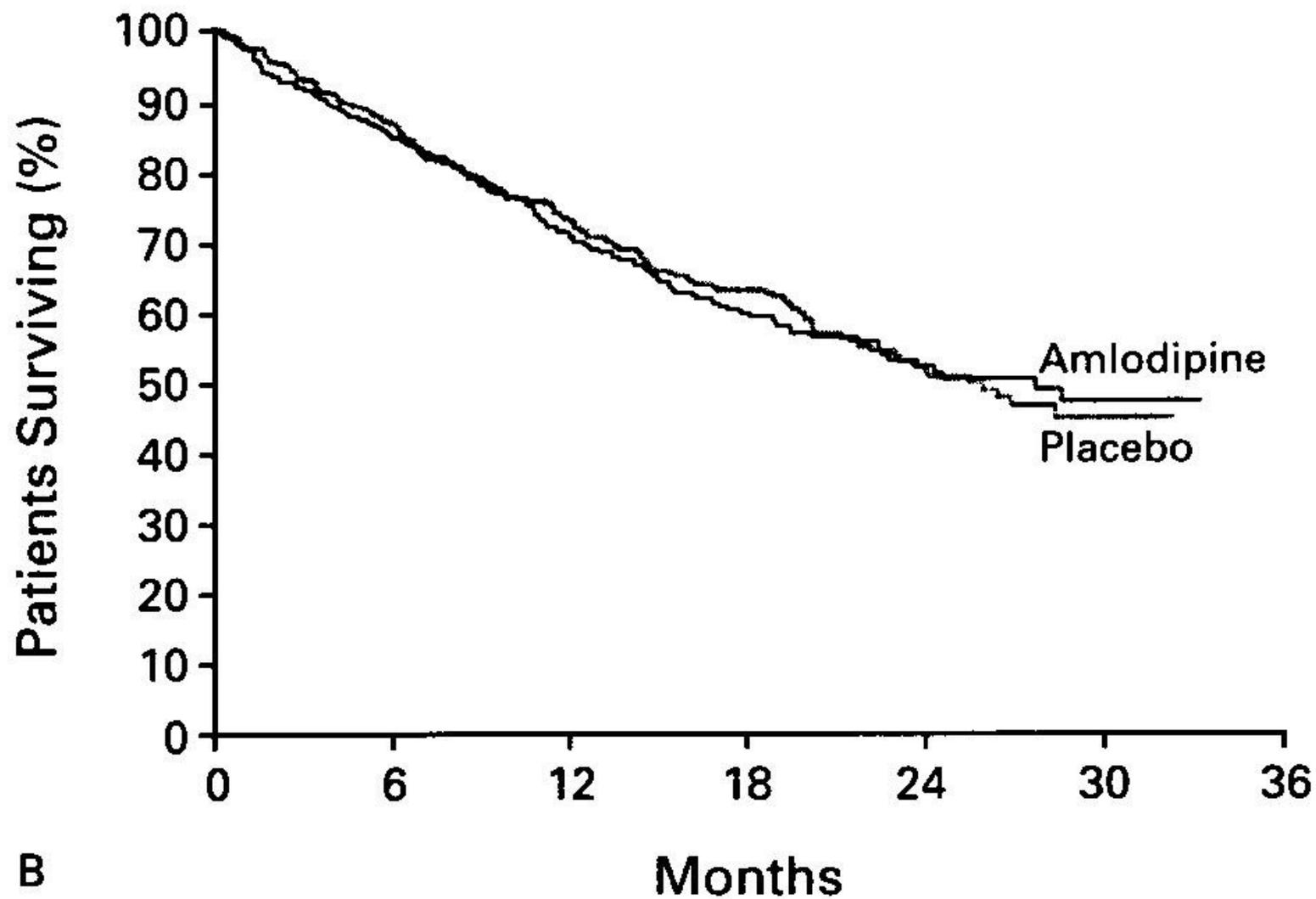
PRAISE I (P=0.07)



PRAISE I - Interaction

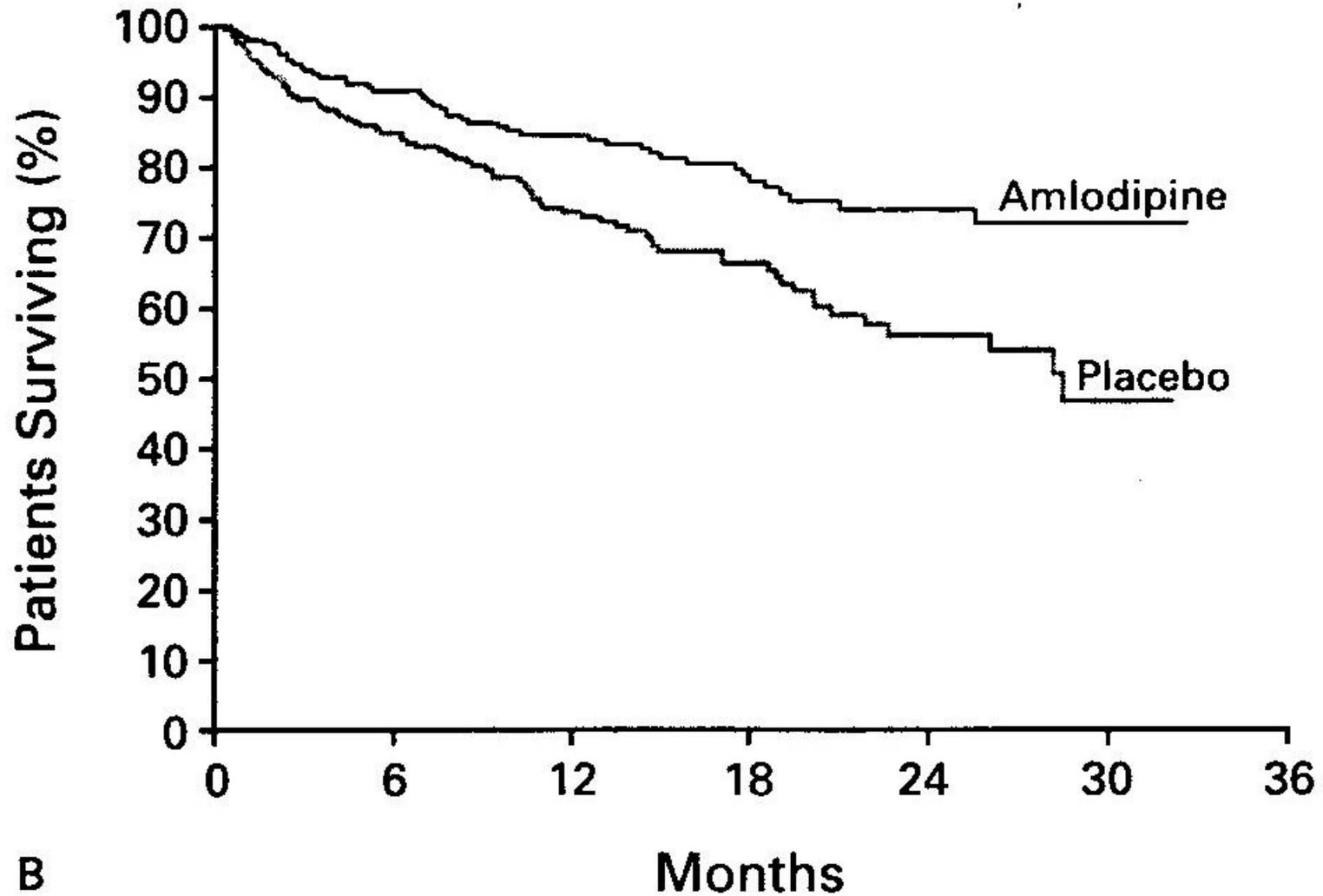
- Overall $P = 0.07$
- Etiology by Treatment Interaction
 $P = 0.004$
- So, break analysis down by subgroups
 - Ischemic Subgroup $P = \text{NS}$
 - Non-Ischemic subgroup $P < 0.001$
- By conventional statistical procedures, we might declare the non-ischemic group a success

PRAISE I - Ischemic



B

PRAISE I – Non- Ischemic



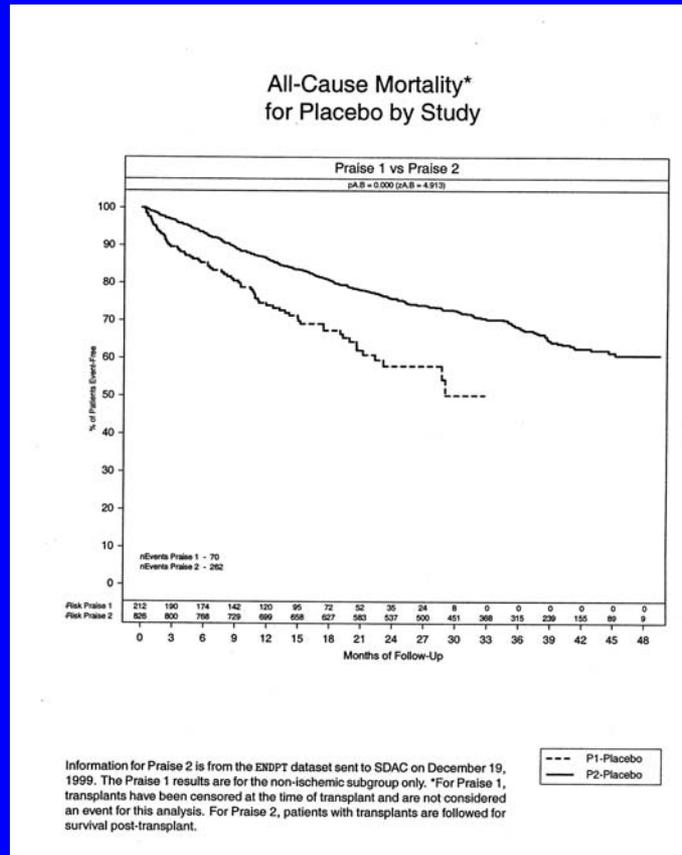
B

PRAISE II

- Investigators repeated PRAISE I for non-ischemic strata
 - Amlodipine vs. placebo
 - Randomized double-blind
 - 1653 patients
 - Mortality the primary outcome
- Final results: $RR \cong 1.0$
- Despite significant interaction in PRAISE I, repeat of subgroup failed to confirm

PRAISE I vs PRAISE II

Placebo arms



Event Classification

- Cause specific events sometimes used to focus on likely treatment effect
- Definitions must be made in advance of classification
- Classification process must be blinded to intervention; otherwise potential bias
- Two separate classification committees might not agree

Anturane Reinfarction Trial

Sudden Death (NEJM, 1980)

Category	Source	Placebo	Anturane	P-value
All patients & all sudden deaths	<i>NEJM</i>	48/817	30/812	0.03
	AC	39/817	28/812	0.17
"Eligible" patients & all sudden deaths	<i>NEJM</i>	46/785	28/775	0.03
	AC	37/782	25/773	0.12

- Problem of cause specific definitions
- AC = Another review committee

Time Dependent Covariate Adjustment

- Classic covariate adjustment uses baseline prognostic factors only
 - Adjust for Imbalance
 - Gain Efficiency
- Adjustment by time dependent variables not recommended in clinical trials (despite Cox time dependent regression model)
- Habit from epidemiology studies

Coronary Drug Project

5-Year Mortality

Example

Baseline Cholesterol	Cholesterol Change	% Deaths	
		Clofibrate	Placebo
< 250mg%*	Fall	16.0	21.2
< 250	Rise	25.5	18.7
≥ 250 mg%	Fall	18.1	20.2
> 250 **	Rise	15.5	21.3

- Little change in placebo group
- Best to have
 - a. Low cholesterol getting lower *
 - b. High cholesterol getting higher **

Composite Endpoint Rationale

- May reduce Sample Size by increasing event rates
 - Assumes each component sensitive to intervention
 - Otherwise, power can be lost
- Death + x + y avoids a competing risk problem
 - Death is a competing risk to all other morbid events, probably not independent
 - Can't look at x or y alone

Problems with Composite Outcomes

- Relevance of a mixed set of components
 - Adding softer outcomes
- Adding irrelevant components could cause a loss of power
- Failure to ascertain components
- Interpretability if individual components go in different directions
 - e.g. WHI global index– Overall, the same
 - Death: similar
 - Fractures: positive
 - DVTs, PEs: negative

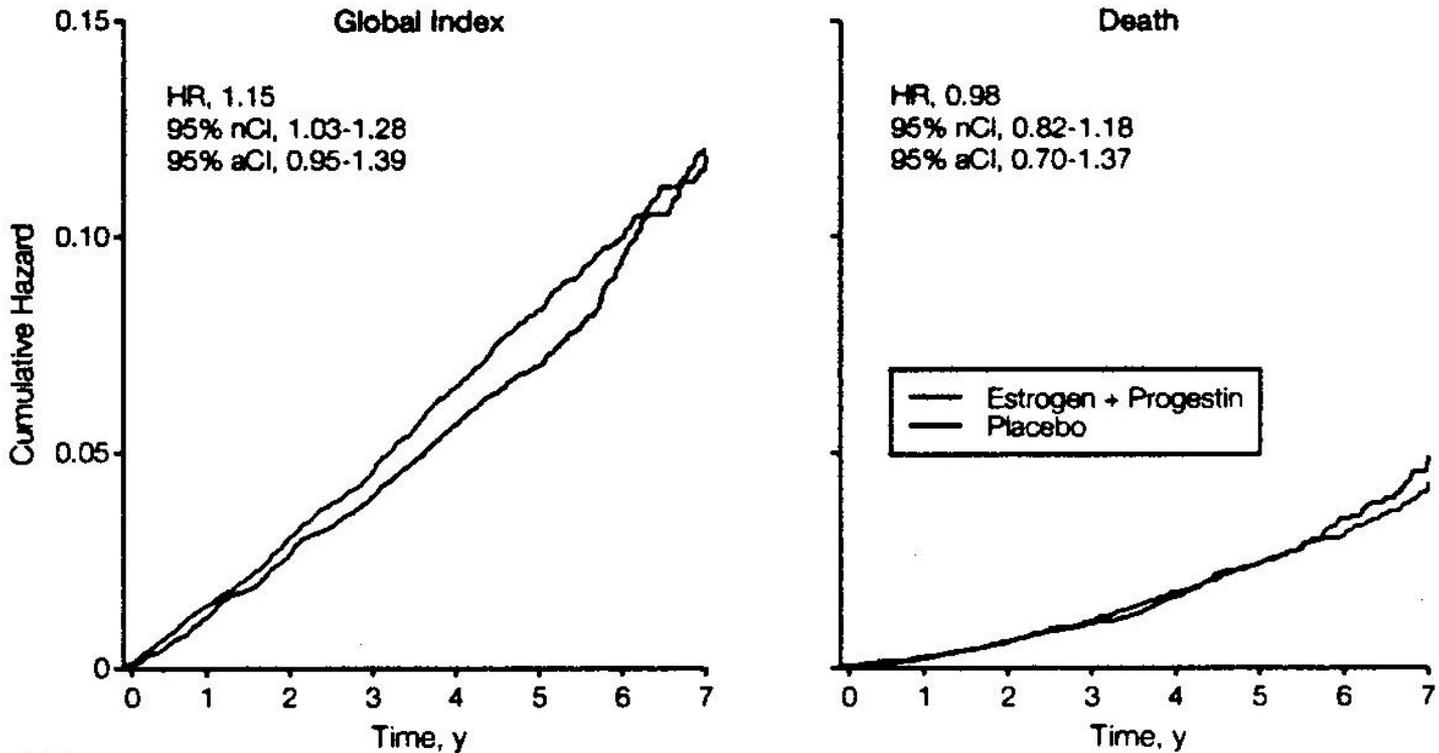
WOMEN'S HEALTH INITIATIVE

JAMA 288(3):321-33, 2002

- A large factorial trial evaluating HRT, low fat diet and calcium
- Multiple outcomes for each treatment
- For HRT
 - Coronary heart disease (MI & CHD death)
 - Invasive breast cancer
 - Fractures
 - Global index (death, CHD, stroke, PE, breast cancer, hip fracture)

WHI

Kaplan-Meier Estimates of Cumulative Hazards for Global Index and Death



No. at Risk

Estrogen +

Progestin 8506 8291 8113 7927 6755 4058 1964 758

Placebo 8102 7939 7774 7607 6425 3794 1662 495

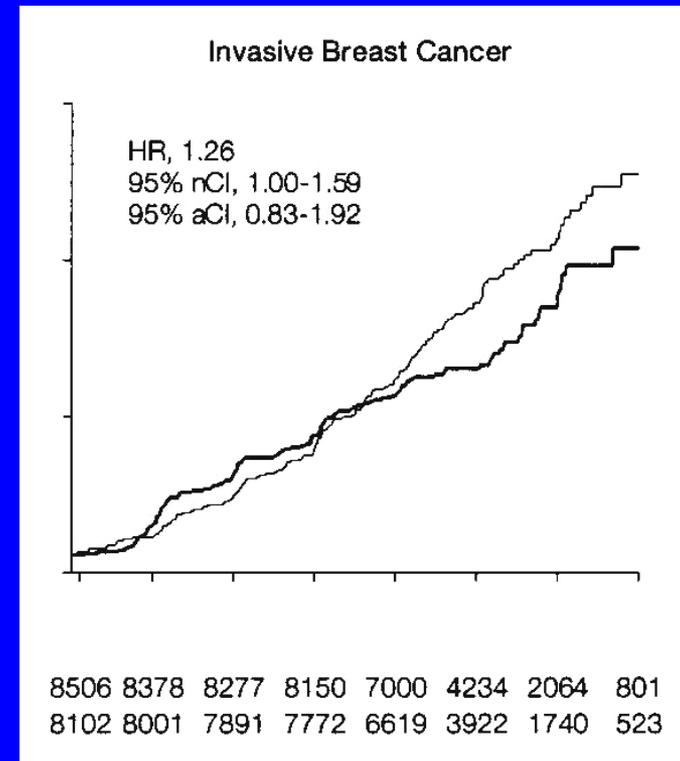
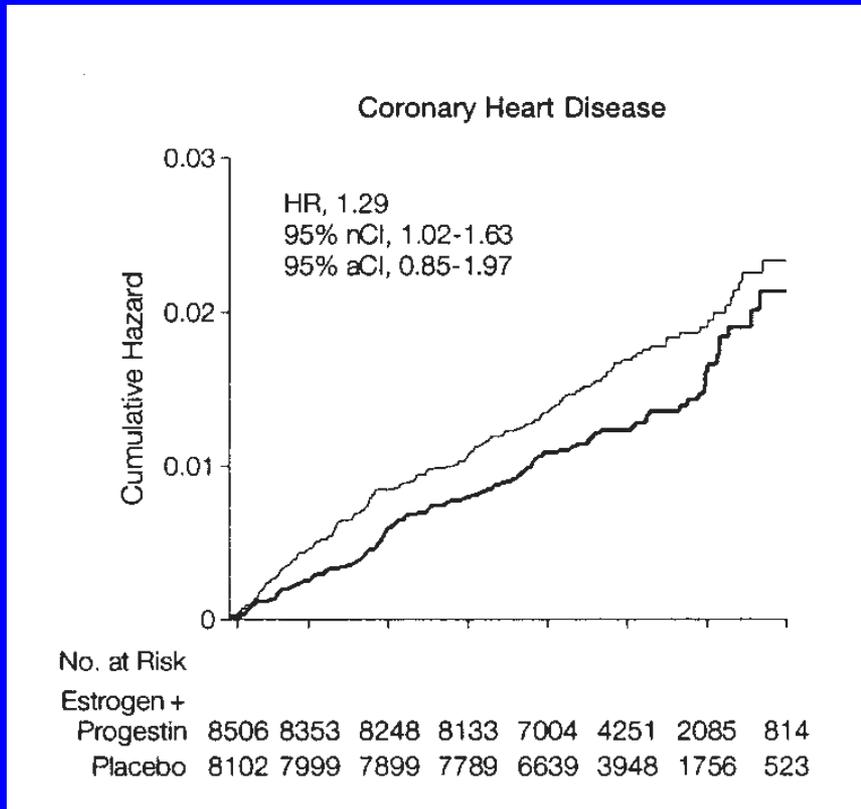
8506 8388 8313 8214 7095 4320 2121 828

8102 8018 7936 7840 6697 3985 1777 530

HR indicates hazard ratio; nCI, nominal confidence interval; and aCI, adjusted confidence interval.

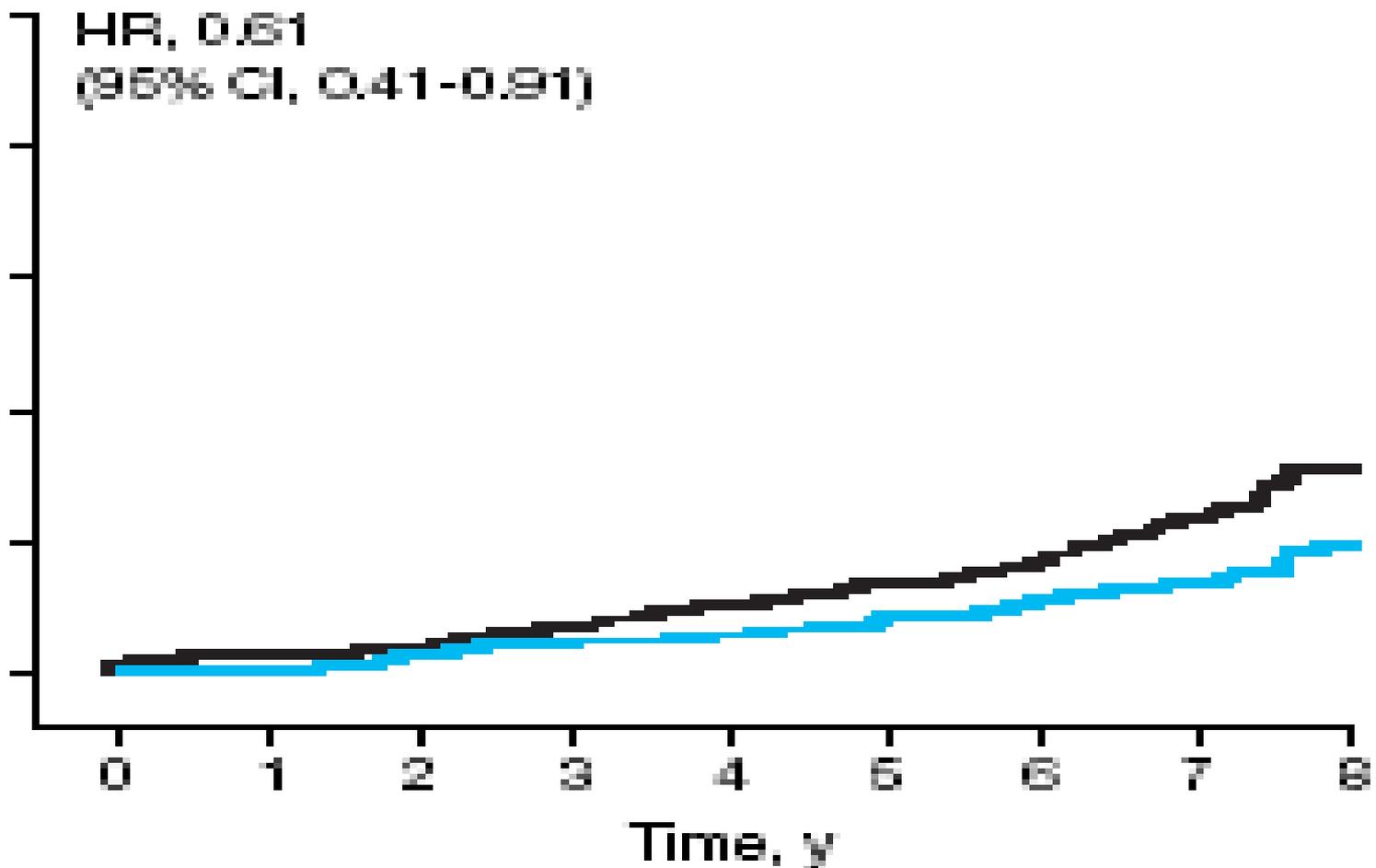
WHI

Kaplan-Meier Estimates of Cumulative Hazards for Selected Clinical Outcomes



WHI-E Hip Fracture

Hip Fracture



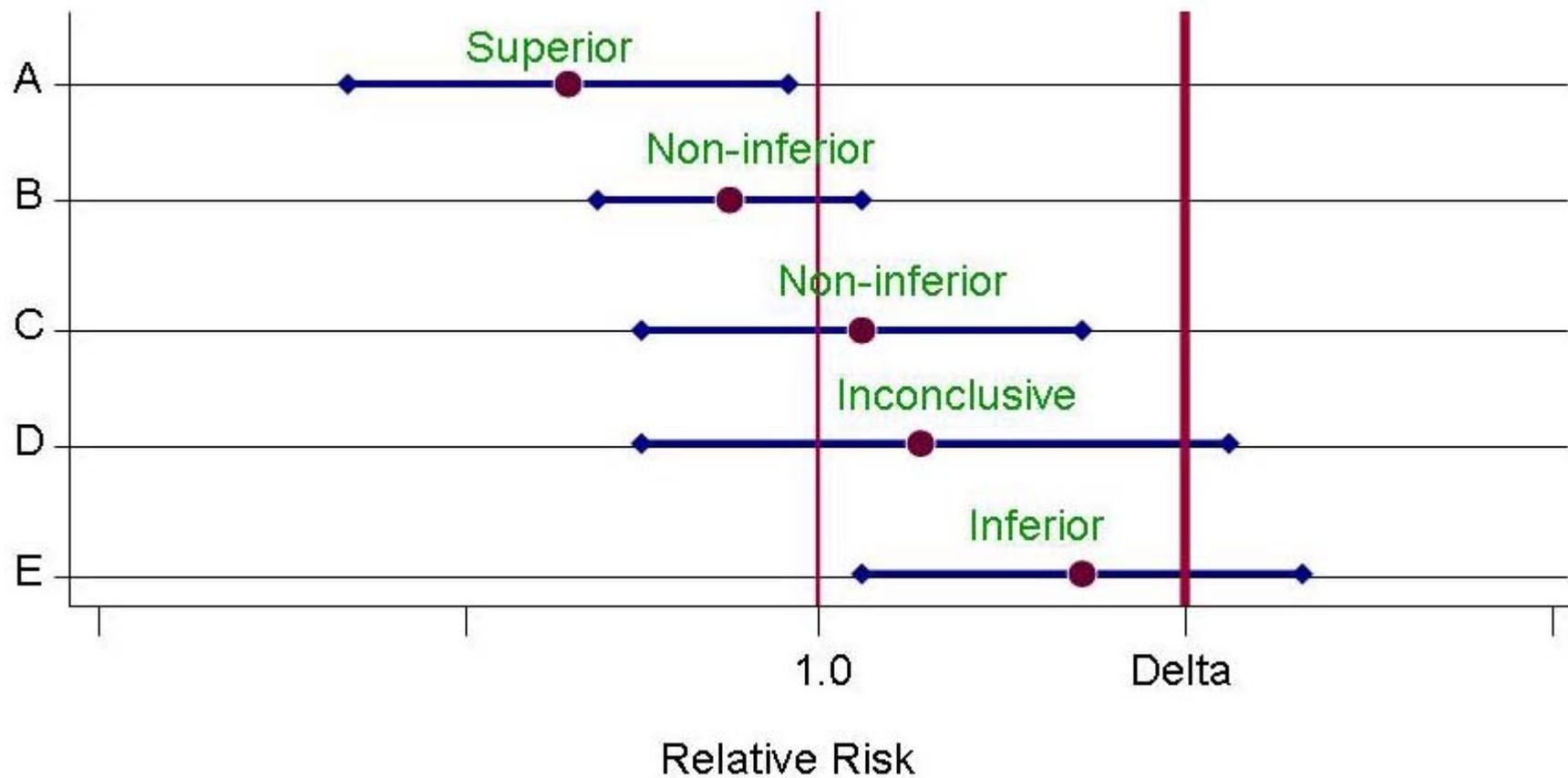
Superiority Trial Design

- Groups
 - T = New experimental intervention
 - C = Control or standard intervention
 - P = Placebo
- Trial designs
 - $T > P$
 - $T > C$
 - $T + C > C$

Non-Inferiority Trials

- Design
 - $T-C < \delta$
 - δ = a predefined margin of indifference
- Must pre-specify margin δ
- Outcome measure & δ
 - Absolute difference
 - Relative difference
- Control must be effective; best available
- Need outstanding compliance

Trial Design



Challenges for Non Inferiority Designs

- Different goals than superiority trials
- Challenges in the design
- Challenges in their conduct
- Challenges in their analyses
- Despite the challenges, probably have to learn to live with them
- Not there yet, in my opinion

Summary

- A well designed & executed trial may be invalidated by issues in the analysis
- Must always include ITT; may add other analyses
- Censored follow-up (except for end of study) and missing data need to be minimized
- Subgroups used cautiously, based on baseline covariates, not post randomization variables such as compliance, biomarker change
- NI designs are an extreme challenge to design and analyze