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Patient Withdrawn in Analysis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

<table>
<thead>
<tr>
<th></th>
<th>Anturane</th>
<th>Placebo</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Randomized</td>
<td>813</td>
<td>816</td>
<td>1629</td>
</tr>
<tr>
<td>Ineligible</td>
<td>38</td>
<td>33</td>
<td>71</td>
</tr>
</tbody>
</table>

- 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

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

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)

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

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

Method I

%Drug   #Pts.    Method I = \frac{\text{dose received}}{\text{total protocol dose}}

\begin{align*}
\geq 85 & \quad \bullet \quad 94 & \\
65-84 & \quad \triangle \quad 13 & \\
<65 & \quad \circ \quad 62 & 
\end{align*}

Method II

%Drug   #Pts.    Method II = \frac{\text{dose received}}{\text{dose while on study}}

\begin{align*}
\geq 85 & \quad \bullet \quad 139 & \\
65-84 & \quad \triangle \quad 14 & \\
<65 & \quad \circ \quad 16 & 
\end{align*}
C: Off Drug ≠ 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
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
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:

- **OPT Alone**
  - Optimal Pharmacological Therapy (OPT)
- **OPT + CRT**
  - (OPT) + CRT (CONTAK TR®/EASYTRAK®)
- **OPT + CRT-D**
  - (OPT) + CRT + ICD (CONTAK CD®/EASYTRAK®)

Randomization stratifications:
by site, +/- β-blocker therapy

Target Time to Implant ≤ 2 days from randomization

HFSA Late-Breaker Sept 24, 2003
• 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

HFSA Late-Breaker Sept 24, 2003
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

CRT vs. OPT: RR = 20%, p=0.008 (Critical boundary=0.014)
CRT-D vs. OPT: RR = 20%, p=0.007 (Critical boundary=0.022)

% of Patients Event-Free

12-month Event Rates
OPT: 68%
CRT: 55% (AR=13%)
CRT-D: 56% (AR=12%)

Days from Randomization
COMPANION: Secondary Endpoint of All-Cause Mortality

CRT vs. OPT: RR = 24%, p=0.060 (Critical boundary=0.014)
CRT-D vs. OPT: RR = 36%, p=0.003 (Critical boundary=0.022)

% of Patients Event-Free

OLT: 19%
CRT: 15% (AR=4%)
CRT-D: 12% (AR=7%)

12-month Event Rates

Days from Randomization

HFEA Late-Breaker September 24, 2003
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

HFSA Late-Breaker Sept 24, 2003
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

Placebo
\( p = 0.0062 \) (adjusted)
\( p = 0.00009 \) (nominal)
Metoprolol CR/XL

Risk reduction = 34%
<table>
<thead>
<tr>
<th>Total mortality</th>
<th>Number of deaths</th>
<th>Metoprolol CR/XL better</th>
<th>Placebo better</th>
</tr>
</thead>
<tbody>
<tr>
<td>NYHA II</td>
<td>103</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NYHA III</td>
<td>232</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NYHA IV</td>
<td>27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EF lower tertile</td>
<td>186</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EF middle + upper tertile</td>
<td>176</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ischaemic aetiology</td>
<td>264</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-ischaemic aetiology</td>
<td>98</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-smoker</td>
<td>306</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smoker</td>
<td>56</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age upper tertile</td>
<td>153</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age middle + lower tertile</td>
<td>209</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>298</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>64</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Previous MI</td>
<td>196</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No previous MI</td>
<td>166</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diabetes mellitus</td>
<td>110</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No diabetes mellitus</td>
<td>252</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Previous hypertension</td>
<td>158</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No previous hypertension</td>
<td>204</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HR lower tertile</td>
<td>126</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HR middle + upper tertile</td>
<td>236</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Systolic BP lower tertile</td>
<td>172</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Systolic BP middle + upper tertile</td>
<td>190</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diastolic BP lower tertile</td>
<td>144</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diastolic BP middle + upper tertile</td>
<td>217</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Relative risk (95% CI)
### All Patients Randomized

#### Total Mortality

<table>
<thead>
<tr>
<th>Country</th>
<th>Favors Meto CR/XL</th>
<th>Favors Placebo</th>
<th>No. of deaths Meto CR/XL/Placebo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belgium</td>
<td>3/13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Czech Republic</td>
<td>9/17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Denmark</td>
<td>11/11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Finland</td>
<td>0/2</td>
<td></td>
<td>Not estimable</td>
</tr>
<tr>
<td>Germany</td>
<td>19/31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hungary</td>
<td>16/29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Iceland</td>
<td>2/2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Norway</td>
<td>6/11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poland</td>
<td>8/8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sweden</td>
<td>2/9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Switzerland</td>
<td>0/1</td>
<td></td>
<td>Not estimable</td>
</tr>
<tr>
<td>The Netherlands</td>
<td>14/25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UK</td>
<td>4/9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>USA</td>
<td>51/49</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All countries</td>
<td>145/217</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Total Mortality/Any Hosp.

<table>
<thead>
<tr>
<th>Country</th>
<th>Favors Meto CR/XL</th>
<th>Favors Placebo</th>
<th>No. of events Meto CR/XL/Placebo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belgium</td>
<td>31/31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Czech Republic</td>
<td>35/50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Denmark</td>
<td>58/60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Finland</td>
<td>6/4</td>
<td></td>
<td>Not estimable</td>
</tr>
<tr>
<td>Germany</td>
<td>88/100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hungary</td>
<td>57/72</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Iceland</td>
<td>6/10</td>
<td></td>
<td>7.3</td>
</tr>
<tr>
<td>Norway</td>
<td>41/48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poland</td>
<td>26/25</td>
<td></td>
<td>2.6</td>
</tr>
<tr>
<td>Sweden</td>
<td>15/27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Switzerland</td>
<td>5/4</td>
<td></td>
<td>Not estimable</td>
</tr>
<tr>
<td>The Netherlands</td>
<td>63/91</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UK</td>
<td>26/29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>USA</td>
<td>184/216</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All countries</td>
<td>641/767</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Total Mortality/CHF Hosp.

<table>
<thead>
<tr>
<th>Country</th>
<th>Favors Meto CR/XL</th>
<th>Favors Placebo</th>
<th>No. of events Meto CR/XL/Placebo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belgium</td>
<td>13/21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Czech Republic</td>
<td>25/36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Denmark</td>
<td>24/28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Finland</td>
<td>2/3</td>
<td></td>
<td>Not estimable</td>
</tr>
<tr>
<td>Germany</td>
<td>44/63</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hungary</td>
<td>31/48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Iceland</td>
<td>3/4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Norway</td>
<td>17/28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poland</td>
<td>16/18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sweden</td>
<td>5/14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Switzerland</td>
<td>1/3</td>
<td></td>
<td>Not estimable</td>
</tr>
<tr>
<td>The Netherlands</td>
<td>28/52</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UK</td>
<td>11/12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>USA</td>
<td>91/109</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All countries</td>
<td>311/439</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Relative risk and 95% confidence interval
BETA-BLOCKER HF TRIALS

CIBIS-2 1998

Survival

0.8

0.6

0.4

0.2

0

Time after inclusion (days)

Bisoprolol

Placebo

p<0.0001

MERIT-HF 1999

Cumulative mortality (%)

p=0.00062 (adjusted)

p=0.0009 (nominal)

COPERNICUS 2001

Survival (% of patients)

100

90

80

70

60

50

40

30

20

10

0

0 3 6 9 12 15 18 21

Months

Carvedilol

Placebo
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)

[Graph showing survival rates over months for Amlodipine and Placebo groups]
PRAISE I - Interaction

- Overall P = 0.07

- Etiology by Treatment Interaction
  P = 0.004

- So, break analysis down by subgroups
  - Ischemic Subgroup P = NS
  - Non-Ischemic subgroup P < 0.001

- By conventional statistical procedures, we might declare the non-ischemic group a success
PRAISE I – Non-Ischemic

![Graph showing survival rates over months for Amlodipine and Placebo groups](image)
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 \( \approx 1.0 \)
• Despite significant interaction in PRAISE I, repeat of subgroup failed to confirm
PRAISE I vs PRAISE II
Placebo arms

All-Cause Mortality
for Placebo by Study

Information for PRAISE II is from the 1999 dataset sent to SDMC on December 19, 1999. The PRAISE I results are for the non-electronic subgroup only. "For PRAISE I, transplants have been censored in the 36 months pre-transplant and are not considered an event for this analysis. For PRAISE II, patients with transplants are followed for normal post-transplant."
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)**

<table>
<thead>
<tr>
<th>Category</th>
<th>Source</th>
<th>Placebo</th>
<th>Anturane</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>All patients &amp; all sudden deaths</td>
<td>NEJM</td>
<td>48/817</td>
<td>30/812</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>AC</td>
<td>39/817</td>
<td>28/812</td>
<td>0.17</td>
</tr>
<tr>
<td>&quot;Eligible&quot; patients &amp; all sudden deaths</td>
<td>NEJM</td>
<td>46/785</td>
<td>28/775</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>AC</td>
<td>37/782</td>
<td>25/773</td>
<td>0.12</td>
</tr>
</tbody>
</table>

- 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**

<table>
<thead>
<tr>
<th>Baseline Cholesterol</th>
<th>Cholesterol Change</th>
<th>% Deaths</th>
<th>Clofibrate</th>
<th>Placebo</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 250mg%*</td>
<td>Fall</td>
<td>16.0</td>
<td>21.2</td>
<td></td>
</tr>
<tr>
<td>&lt; 250</td>
<td>Rise</td>
<td>25.5</td>
<td>18.7</td>
<td></td>
</tr>
<tr>
<td>&gt; 250 mg%</td>
<td>Fall</td>
<td>18.1</td>
<td>20.2</td>
<td></td>
</tr>
<tr>
<td>&gt; 250 **</td>
<td>Rise</td>
<td>15.5</td>
<td>21.3</td>
<td></td>
</tr>
</tbody>
</table>

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

HR, 1.15
95% nCl, 1.03-1.28
95% aCl, 0.95-1.39

HR, 0.98
95% nCl, 0.82-1.18
95% aCl, 0.70-1.37

No. at Risk
Estrogen + Progestin 8506 8291 8113 7927 6755 4058 1964 758
Placebo 8102 7939 7774 7607 6425 3794 1662 495

HR indicates hazard ratio; nCl, nominal confidence interval; and aCl, adjusted confidence interval.

JAMA, 2002
Kaplan-Meier Estimates of Cumulative Hazards for Selected Clinical Outcomes

Coronary Heart Disease

HR, 1.29
95% nCl, 1.02-1.63
95% aCl, 0.85-1.97

Invasive Breast Cancer

HR, 1.26
95% nCl, 1.00-1.59
95% aCl, 0.83-1.92

JAMA, 2002
WHI-E Hip Fracture

HR, 0.61
(95% CI, 0.41-0.91)
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 < $\delta$
  - $\delta$ = a predefined margin of indifference
- Must pre-specify margin $\delta$
- Outcome measure & $\delta$
  - Absolute difference
  - Relative difference
- Control must be effective; best available
- Need outstanding compliance
Trial Design

Relative Risk

- Superior
- Non-inferior
- Inconclusive
- Inferior

Delta
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 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