# Interim Analysis and Data Monitoring in Clinical Trials

Jen-pei Liu, Ph.D. Professor
Division of Biometrics, Department of Agronomy, National Taiwan National University and
Division of Biostatistics
National Health Research Institutes

# Interim analyses and Data Monitoring

- Limited and Finite Resource
  - Ethical
  - Scientific
  - Cost
    - Patients
    - $$$
    - Time

  Optimization of available resource.

# Information

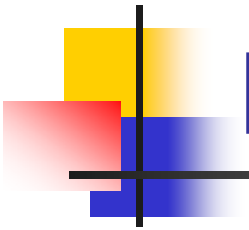
- Sample mean of N observations
- The variance of sample mean $=\sigma^2/N$.
- The information about the population mean provided by the sample mean is $N/\sigma^2$.
- If $\sigma^2 = 1$, the information about the population mean provided by the sample of N observation is simply the sample size N.

# Information

- This is the definition of statistical information which can be also interpreted as the clinical information
- The planned sample size is the minimum clinical and statistical information required to achieve the desired power for detecting a minimum clinically meaningful difference at a pre-determined risk of type I error.

# Maximum information trial

- A clinical trial is allowed to continue until all N subjects, the pre-determined sample size, have completed the scheduled follow-up, then it is referred to as the maximum information trial.
- For a maximum information trial, the total duration is a random variable.



# Maximum duration trial

- If a trial is terminated at the maximum duration, then it is referred to as the maximum duration trial.
- For a maximum duration trial, the total information accumulated in the trial is a random variable.

# Information Time vs. Calendar Time

- $N(T_c)$ is the maximum information obtained at the maximum duration $T_c$.
- $n(t)$ is the number of subjects who complete the study at calendar time t.
- Information time vs. calendar time

$$s(t) = n(t)/N(T_c),\ 0 \le t \le T_c.$$

$t_k$ is the calendar time at which the kth subject completes the study, $k = 0,1,\ldots N$

$$s_k = s(t_k) = n(t_k)/N(T_c),\ 0 \le t \le T_c.$$

$$s(t) = s_k, \text{ if } t \in [t_k, t_{k+1})$$

# Information Time vs. Calendar Time

| Time of DMCM | Calendar Time | Cum. Time | Cum. Death | D=628 Information Time | D=400 Information Time |
|---|---|---|---|---|---|
| 6/1978 | 0 Months | 0 | 0 | 0 | 0 |
| 5/1979 | 11 Months | 0.23 | 56 | 0.09 | 0.14 |
| 10/1979 | 16 Months | 0.33 | 77 | 0.12 | 0.19 |
| 3/1980 | 21 Months | 0.44 | 126 | 0.20 | 0.32 |
| 10/1980 | 28 Months | 0.58 | 177 | 0.28 | 0.44 |
| 4/1981 | 34 Months | 0.71 | 247 | 0.39 | 0.62 |
| 10/1981 | 40 Months | 0.83 | 318 | 0.51 | 0.80 |
| 6/1982 | 48 Months | 1.00 | 400/628 | 1.00 | 1.00 |

# Interim Analyses

Group Sequential Procedures

- Two extreme Cases
  - Single-stage (Fixed sample)
  - Classical Sequential
- Intermediate Case

  Group sequential or multi-stage trials with some adjustments in type I error probabilities at successive stages.
- WARNING !!!

  Failure to do so

  Actual probability of type I error >>> Nominal probability of type I error

# Group Sequential Procedures

- Trial

  Two parallel groups, double-blinded randomized, FIXED SAMPLE Number of patients is sufficient large

- Clinical endpoints
  - Continuous data
    - Systolic and diastolic blood pressures
    - Serum cholesterol levels
  - Categorical data
    - Response
    - Performance status
  - Censored data
    - Time to death from all causes
    - Time to myocardial infarction

# Group Sequential Procedures

- Z statistics can be computed
- Reject the null hypothesis and conclude there is a statistically significant difference between two groups at 5% nominal level if

  absolute value of Z > 1.96

  I.e., either Z < -1.96 or Z > 1.96

# Repeated Significance Test on Accumulated Data

| The Number of Repeated Test at the 5% level | Overall Significance level |
|---|---|
| 1 | 0.05 |
| 2 | 0.08 |
| 3 | 0.11 |
| 4 | 0.13 |
| 5 | 0.14 |
| 10 | 0.19 |
| 20 | 0.25 |
| 50 | 0.32 |
| 100 | 0.37 |
| 1000 | 0.53 |
| $\infty$ | 1.00 |

- Repeated testing increases probability of type Ⅰ error
- Must make adjustment for nominal significant level to be conservative

# Group Sequential Procedures

- Idea
  - Compute summary (Z) statistic at each interim analysis, based on additional groups of new patients
  - Compare statistics to a conservative critical value for an overall 0.05 level of type Ⅰ error probability

# Two parallel groups

A total of N patients is planned to recruit

The number of interim analyses is pre-determined in advance in the protocol.

K interim analyses (stages) including the final analysis

N/2 patients per group at each stage, n = N/k

| Stage | Test Drug | Placebo | Information Fraction | Z-statistic |
|---|---|---|---|---|
| 1 | n/2 | n/2 | 1/k(n) | Z1 |
| 2 | 2n/2 | 2n/2 | 2/k(2n) | Z2 |
| 3 | 3n/2 | 3n/2 | 3/k(3n) | Z3 |
| ⋮ | | | | |
| K | kn/2 | kn/2 | 1(kn) | Zk |

At each of the K stages compute the Z-statistics.

# Data Structure of Group Sequential Methods

| Number of Interim analysis (K) | Randomization | | Information Time | Z- Statistic |
|---|---|---|---|---|
| | Test Drug | Placebo | | |
| 1 | n/2 | n/2 | 1/k(n) | Z1 |
| 2 | 2n/2 | 2n/2 | 2/k(2n) | Z2 |
| 3 | 3n/2 | 3n/2 | 3/k(3n) | Z3 |
| | | | | |
| K | kn/2 | kn/2 | 1(kn) | Zk |

*Source*: Reproduced with permission for overall significance levels from Armitage et al. (1969)

# Haybittle-Peto method
## Haybittle (1971), Peto (1976)

- Two parallel groups
- A total of N patients is planned to recruit
- K interim analyses (stages) including the final analysis
- N/2 patients per group at each stage, n = N/k
  - Simple, Ad Hoc
  - For interim analyses, use conservative critical values e.g. ± 3.0
  - For final analysis, no adjustment is required, i.e., use ± 1.96
  - At each stage compute Z-statistic
  - If Z-statistic crosses ±3.0, then reject the null hypothesis and recommend the possibility of terminating the trial.
  - Final use ± 1.96
  - Otherwise, continue the trial to the next stage

# Pocock's method
## Pocock (1977)

- Two parallel groups
- A total of N patients is planned to recruit
- K interim analyses (stages) including the final analysis
- N/2 patients per group at each stage, n = N/k
  - For interim analyses, use a conservative critical values e.g. $\pm C_k$
  - The critical values depend upon the number of planned interim analyses
  - At each stage compute Z-statistic
  - If Z-statistic crosses $\pm C_k$, then reject the null hypothesis and recommend the possibility of terminating the trial.
  - Otherwise, continue the trial to the next stage

# O'Brien-Fleming's Method
## O'Brein and Fleming (1979)
## Fleming, Harrington, and O'Brien (1984)

- Two parallel groups
- A total of N patients is planned to recruit
- K interim analyses (stages) including the final analysis
- N/2 patients per group at each stage, n = N/k
  - For interim analyses, use a conservative critical values e.g. $\pm C_{ik}$
  - The critical values depend upon the number of planned interim analyses and stage of interim analysis
  - At each stage compute Z-statistic
  - If Z-statistic crosses $\pm C_{ik}$, then reject the null hypothesis and recommend the possibility of terminating the trial.
  - Otherwise, continue the trial to the next stage

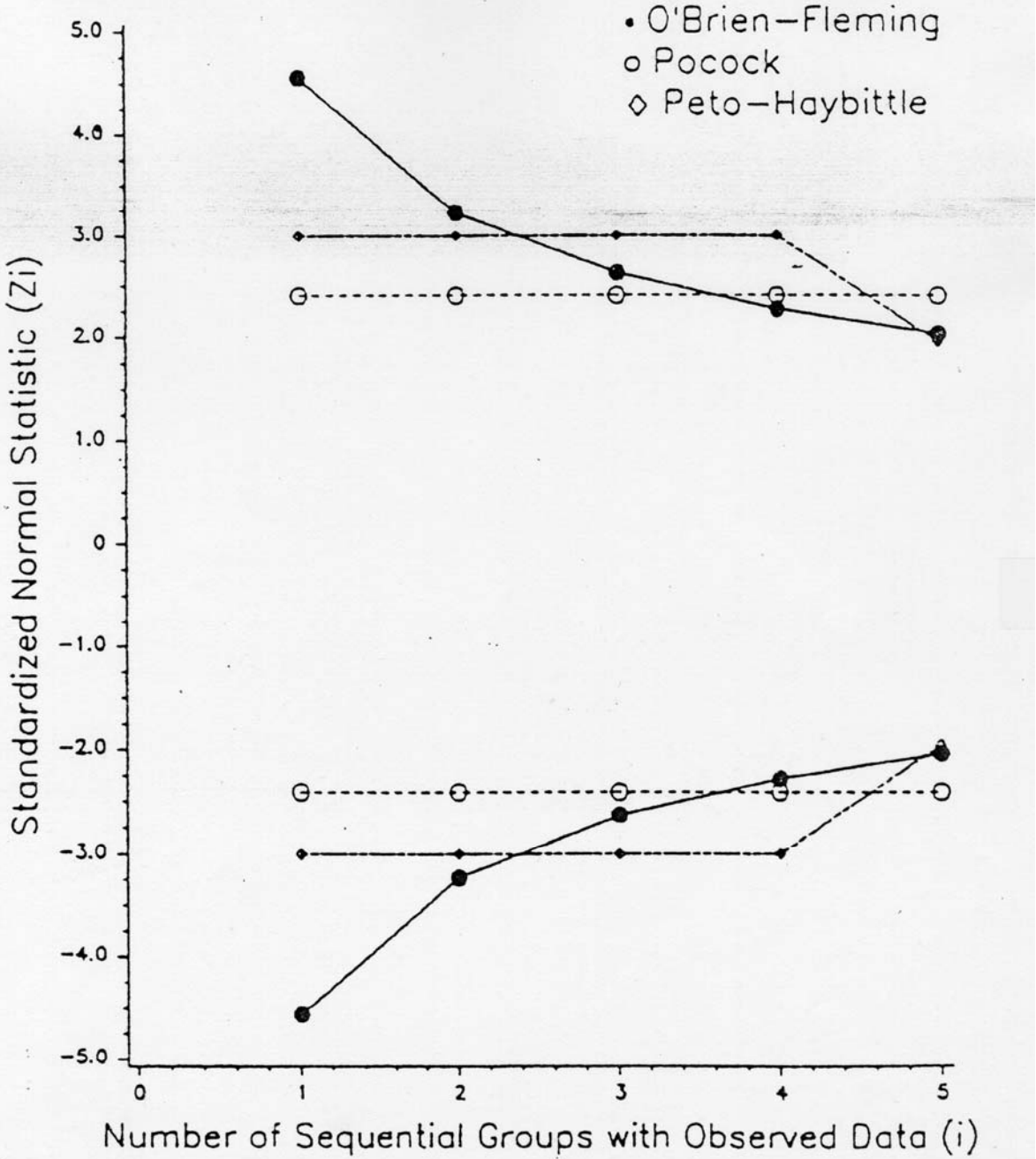



**Figure 9.6.1** Two-sided 5% boundaries for Pocock, O'Brien-Fleming, and Peto-Haybittle methods. (Source: Lan and DeMets, 1994).

# Repeated Confidence Intervals (RCI)

- Jennsion and Turnball (1989)
- Alternative formulation to the interim repeated significance testing
- These confidence intervals use the same critical values as used interim analyses
- Computation is the same as the usual confidence interval except for the critical values

$$d_i \pm c_i SE(d_i)$$

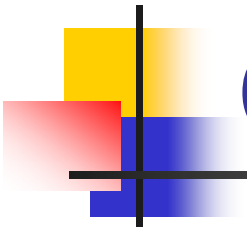
- Offer something more than just p-value

# The Conservative Critical Values

Two-sided: 0.05, One-sided: 0.025

| Number of Pocock | | O'Brien-Fleming | |
|---|---|---|---|
| Stage (K) | Value | Multiplier | Value |
| 1 | 1.960 | sqrt(1/i) | 1.960 |
| 2 | 2.177 | sqrt(2/i) | 1.977 |
| 3 | 2.289 | sqrt(3/i) | 2.005 |
| 4 | 2.362 | sqrt(4/i) | 2.025 |
| 5 | 2.412 | sqrt(5/i) | 2.040 |
| 6 | 2.453 | sqrt(6/i) | 2.052 |
| 7 | 2.486 | sqrt(7/i) | 2.064 |
| 8 | 2.512 | sqrt(8/i) | 2.071 |
| 9 | 2.536 | sqrt(9/i) | 2.081 |
| 10 | 2.555 | sqrt(10/i) | 2.086 |

Modified from Jennison and Turnball (1991)



# Group Sequential Procedures

- Limitations
  - Number of planned interim analyses specified in advance
  - Require equal increment of patients for each stage
  - Limit Data Monitoring Committee
    - Ethical concerns
    - Scientific concerns

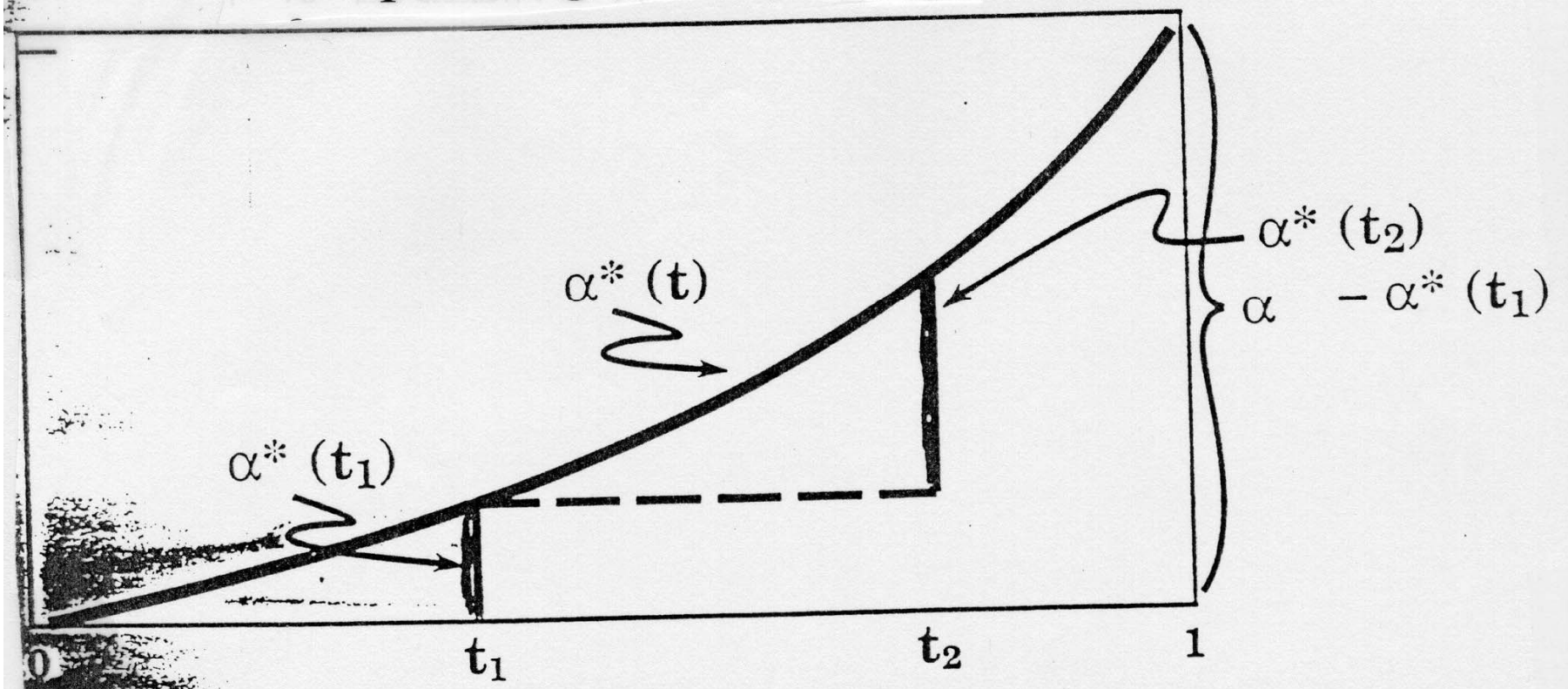
# Spending Functions
## DeMets and Lan (1994)

- Extend previous group sequential procedures to gain more flexibility
- Define a function to spend the overall nominal significance level
- The spending function $\alpha$*(t) is a increasing function of information fraction t

$$\alpha^*(t) = \begin{cases} 0, if\ \ t = 0 \\ \alpha, if\ \ t = 1 \end{cases}$$

- Specify the spending function in advance
- Can not change the spending function during the trial

# The spending function $\alpha^*(t)$




**Find Boundary Values $C_i$**

$C_1 \Rightarrow P\{ Z \geq C_1\} = \alpha^* (t_1)$

$C_2 \Rightarrow P\{ Z_1 < C_1, Z_2 \geq C_2\} = \alpha^* (t_2) - \alpha^* (t_1)$

# Spending Functions
## DeMets and Lan (1994)

- Compute the nominal significance level you need to spend at an interim analysis based on information fraction,

  i.e., $\alpha^*(t_2) - \alpha^*(t_1)$, $t_1 < t_2$
- Compute the corresponding critical values and perform the interim analyses as other group sequential methods
- No need to specify the number of interim analyses
- No need to specify the time to perform interim analyses

# Different Forms of Alpha Spending Functions

| | Approximation |
|---|---|
| $\alpha_1(s)=2\{1-\varphi[z(\alpha/2)/\sqrt{s}]\}$ | O'Brien-Fleming |
| $\alpha_2(s)=\alpha \ln[1+(e-1)s]$ | Pocock |
| $\alpha_3(s)=\alpha s^{\theta},\ \theta>0$ | Lan-DeMets-Kim |
| $\alpha_4(s)=\alpha[(1-e^{-\zeta s})/(1-e^{-\zeta})],\ \zeta \neq 0$ | Hwang-Shih |

# Examples of Boundaries by Alpha Spending Function

| Interim Analysis (s) | O'Brien-Fleming | $\alpha_1(s)$ | pocock | $\alpha_2(s)$ | $\alpha_3(s)[\theta\text{-}1]$ |
|---|---|---|---|---|---|
| 1(0.2) | 4.56 | 4.90 | 2.41 | 2.44 | 2.58 |
| 2(0.4) | 3.23 | 3.35 | 2.41 | 2.43 | 2.49 |
| 3(0.6) | 2.63 | 2.68 | 2.41 | 2.41 | 2.41 |
| 4(0.8) | 2.28 | 2.29 | 2.41 | 2.40 | 2.34 |
| 5(1.0) | 2.04 | 2.03 | 2.41 | 2.39 | 2.28 |

*Note:* Number of interim Analyses =5.

Two-sided: 0.05, One-sided: 0.025

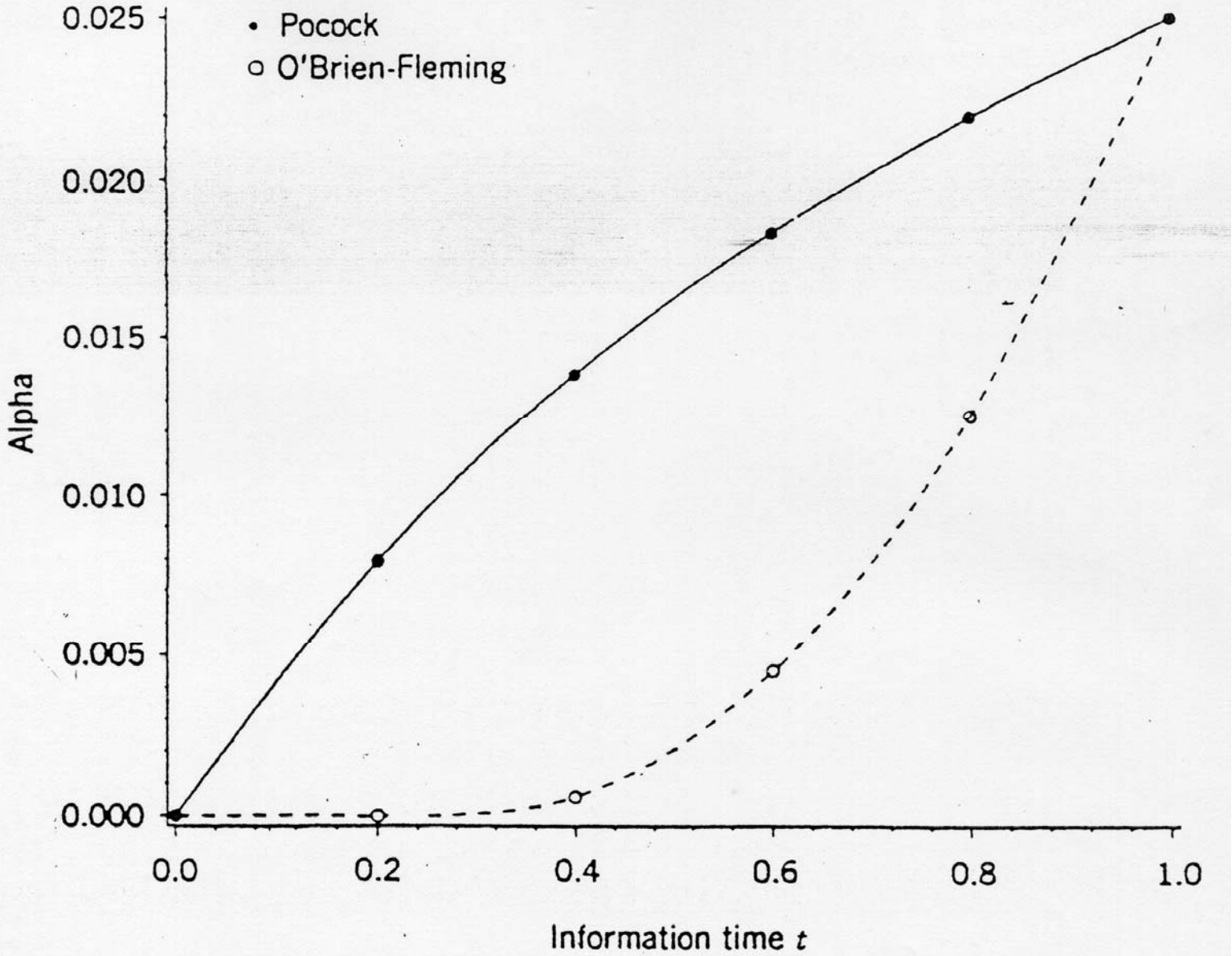



**Figure 9.6.3** One-sided 2.5% alpha spending function for Pocock and O'Brien-Fleming boundaries. (Source: Lan and DeMets, 1994).

# Cumulative Probability of Type I Error

| Interim analysis(s) | Group Sequential Boundaries | | | | | |
|---|---|---|---|---|---|---|
| | Pocock | | | O'Brien-Fleming | | |
| | Value | $\alpha$ (s) | Increment | Value | $\alpha$ (s) | Increment |
| 1(0.2) | 2.41 | 0.0079 | 0.0079 | 4.56 | 0.0000 | $2.6\times10^{-6}$ |
| 2(0.4) | 2.41 | 0.0138 | 0.0059 | 3.23 | 0.0006 | 0.000574 |
| 3(0.6) | 2.41 | 0.0183 | 0.0045 | 2.63 | 0.0045 | 0.0039 |
| 4(0.8) | 2.41 | 0.0219 | 0.0036 | 2.28 | 0.0125 | 0.0080 |
| 5(1.0) | 2.41 | 0.0250 | 0.0031 | 2.04 | 0.0250 | 0.0125 |

# Computation of Spending Probability and Boundaries

$P\{Z(0.2) > 4.56\} = 2.6 \times 10^{-6}$

$P\{Z(0.2) > 4.56 \text{ or } Z(0.4) > 3.23\}$

$= P\{Z(0.2) > 4.56\} + P\{Z(0.2) \le 4.56 \text{ and } Z(0.4) > 3.23\}$

$= 2.6 \times 10^{-6} + 0.000574 = 0.0006$

$P\{Z(0.2) > 4.56 \text{ or } Z(0.4) > 3.23 \text{ or } Z(0.6) > 2.63\}$

$= P\{Z(0.2) > 4.56 \text{ or } Z(0.4) > 3.23\} + P\{Z(0.2) \le 4.56 \text{ and } Z(0.4) \le 3.23 \text{ and } Z(0.6) > 2.63\}$

$= 0.0006 + 0.0039 = 0.0045$

$P\{Z(0.2) > 4.56 \text{ or } Z(0.4) > 3.23 \text{ or } Z(0.6) > 2.63 \text{ or } Z(0.8) > 2.28\}$

$= P\{Z(0.2) > 4.56 \text{ or } Z(0.4) > 3.23 \text{ or } Z(0.6) > 2.63\} +$

$P\{Z(0.2) \le 4.56 \text{ and } Z(0.4) \le 3.23 \text{ and } Z(0.6) \le 2.63 \text{ and } Z(0.8) > 2.28\}$

$= 0.0045 + 0.0080 = 0.0125$

$P\{Z(0.2) > 4.56 \text{ or } Z(0.4) > 3.23 \text{ or } Z(0.6) > 2.63 \text{ or } Z(0.8) > 2.28 \text{ or } Z(1.0) > 2.04\}$

$= P\{Z(0.2) > 4.56 \text{ or } Z(0.4) > 3.23 \text{ or } Z(0.6) > 2.63 \text{ or } Z(0.8) > 2.28\}$

$+ P\{Z(0.2) \le 4.56 \text{ and } Z(0.4) \le 3.23 \text{ and } Z(0.6) \le 2.63 \text{ and } Z(0.8) \le 2.28 \text{ and } Z(1.0) > 2.04\}$

$= 0.0125 + 0.0125 = 0.0250$

# Method of B-values
## Lan and Wittes (1988)

- Objective

  To compute the conditional probability of reaching a statistical significance at the conclusion of the study given the current observed Z-statistic

- Compute $B_t = Z_t \sqrt{t}$
- Compute $q = [1.96 - B_t/t] / \sqrt{1-t}$
- Compute $Pr\{Z>q\}$

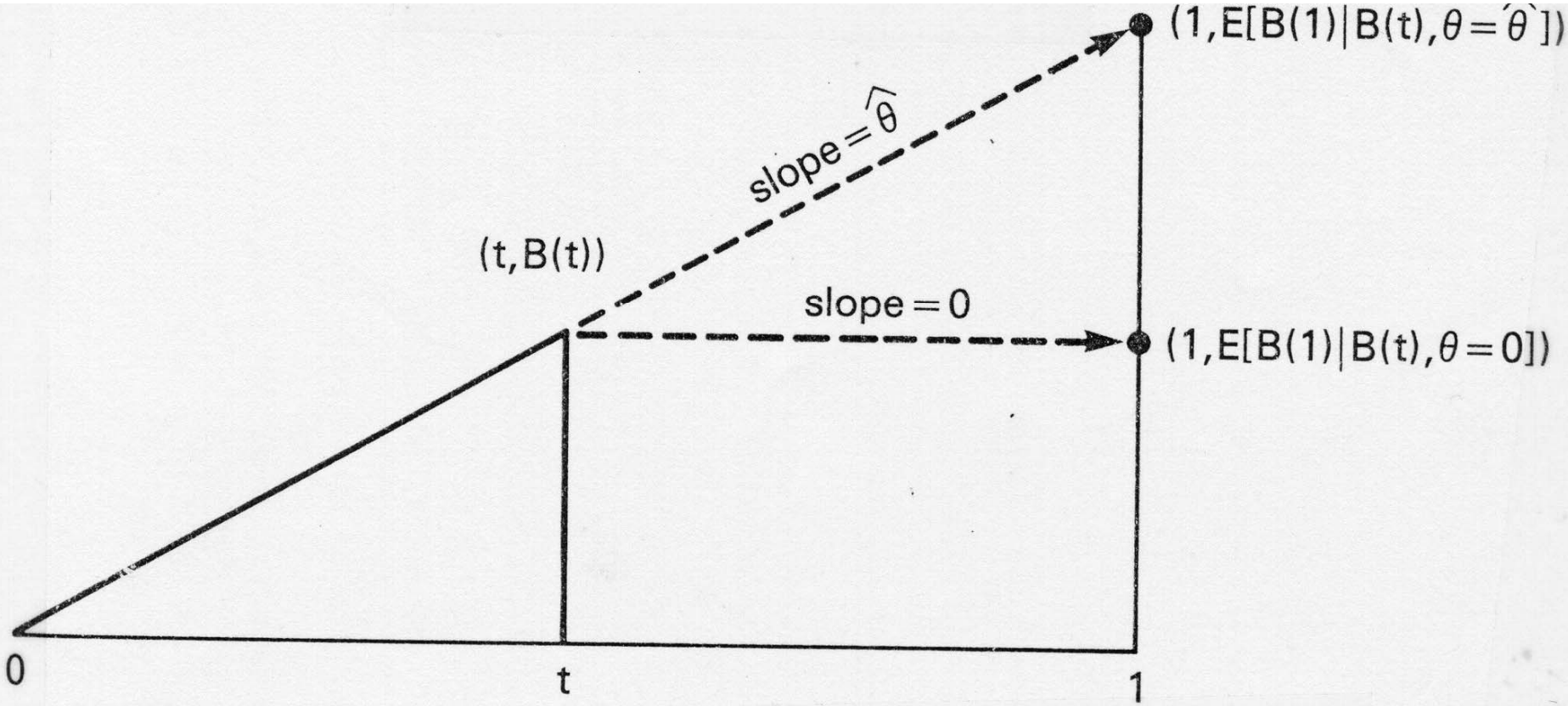



**Figure 1.** Projections of trends.

# Examples

- Beta-Blocker Heart Attack Trial (BHAT)
- Design Features

| | |
|---|---|
| Informed consent | 3,837 patients |
| Randomized | Male and females |
| Double-blinded | 30-69 years of age |
| Placebo-controlled | 5-21 days post M.I. |
| Extended follow-up | Propranolol<br>180 or 240 mg/day |

Preliminary report (1981) JAMA, 246: 2073-2074
Final report (1982) JAMA, 247: 1707-1714

# Beta-Blocker Heart Attack Trial

Baseline Comparisons

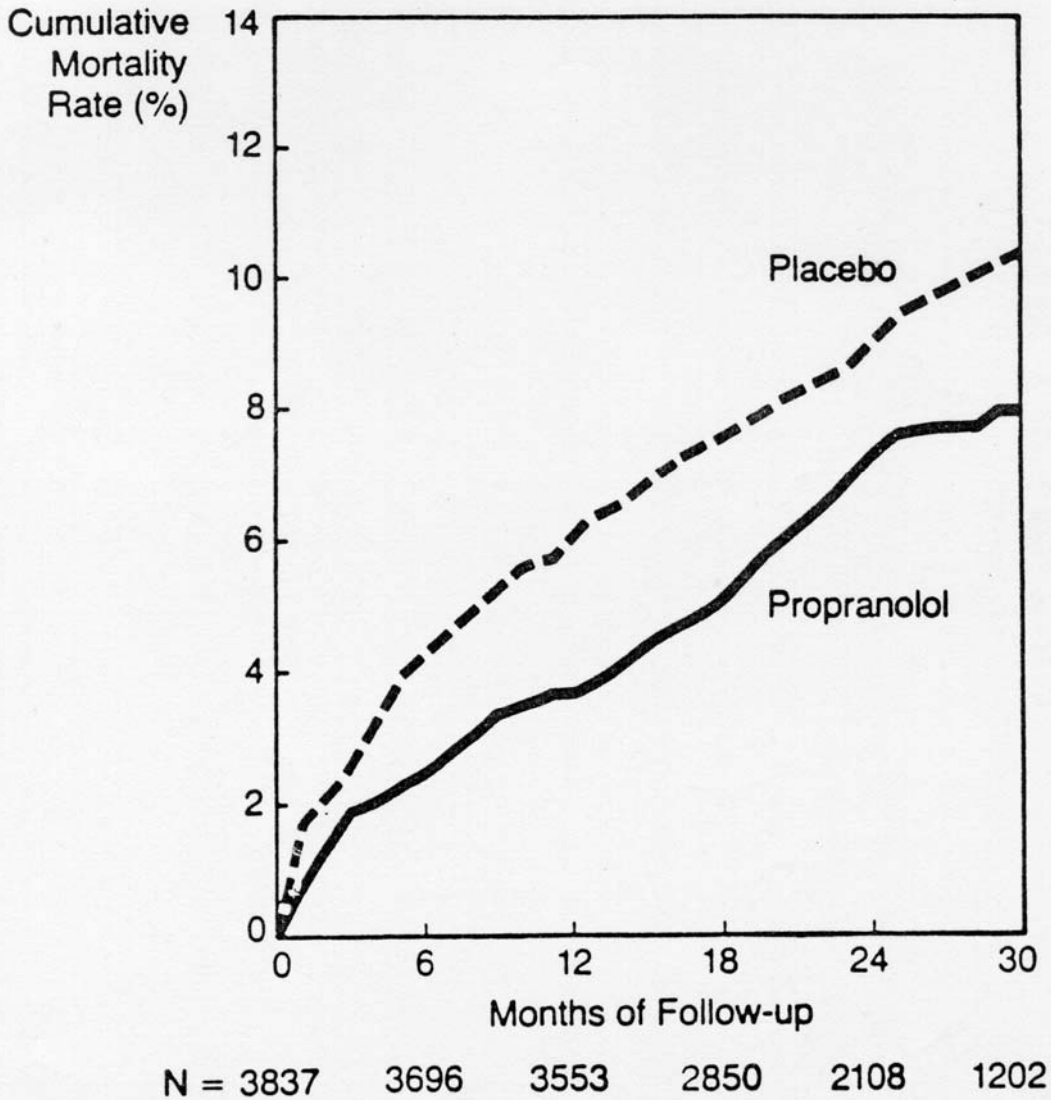
| | Propranolol (N=1,916) | Plcebo (N = 1,921) |
|---|---|---|
| Average Age (yrs.) | 55.2 | 55.4 |
| Male (%) | 83.8 | 85.2 |
| White | 89.3 | 88.4 |
| Systolic B.P. | 112.3 | 111.7 |
| Diastolic B.P. | 72.6 | 72.3 |
| Heart rate | 76.2 | 75.7 |
| Cholesterol | 212.7 | 213.6 |
| Current smoker (%) | 57.3 | 56.8 |

# BHAT Accumulating Survival Data

| Date of DMC Meeting | Propranolol | Placebo | Z (log rank) |
|---|---|---|---|
| May 1979 | 22 / 860 | 34 / 848 | 1.68 |
| October 1979 | 29 / 1080 | 48 / 1080 | 2.24 |
| March 1980 | 50 / 1490 | 76 / 1486 | 2.37 |
| October 1980 | 74 / 1846 | 103 / 1841 | 2.30 |
| April 1981 | 106 / 1916 | 141 / 1921 | 2.34 |
| October 1981 | 135 / 1916 | 183 / 1921 | 2.82* |

* DMC terminated the BHAT

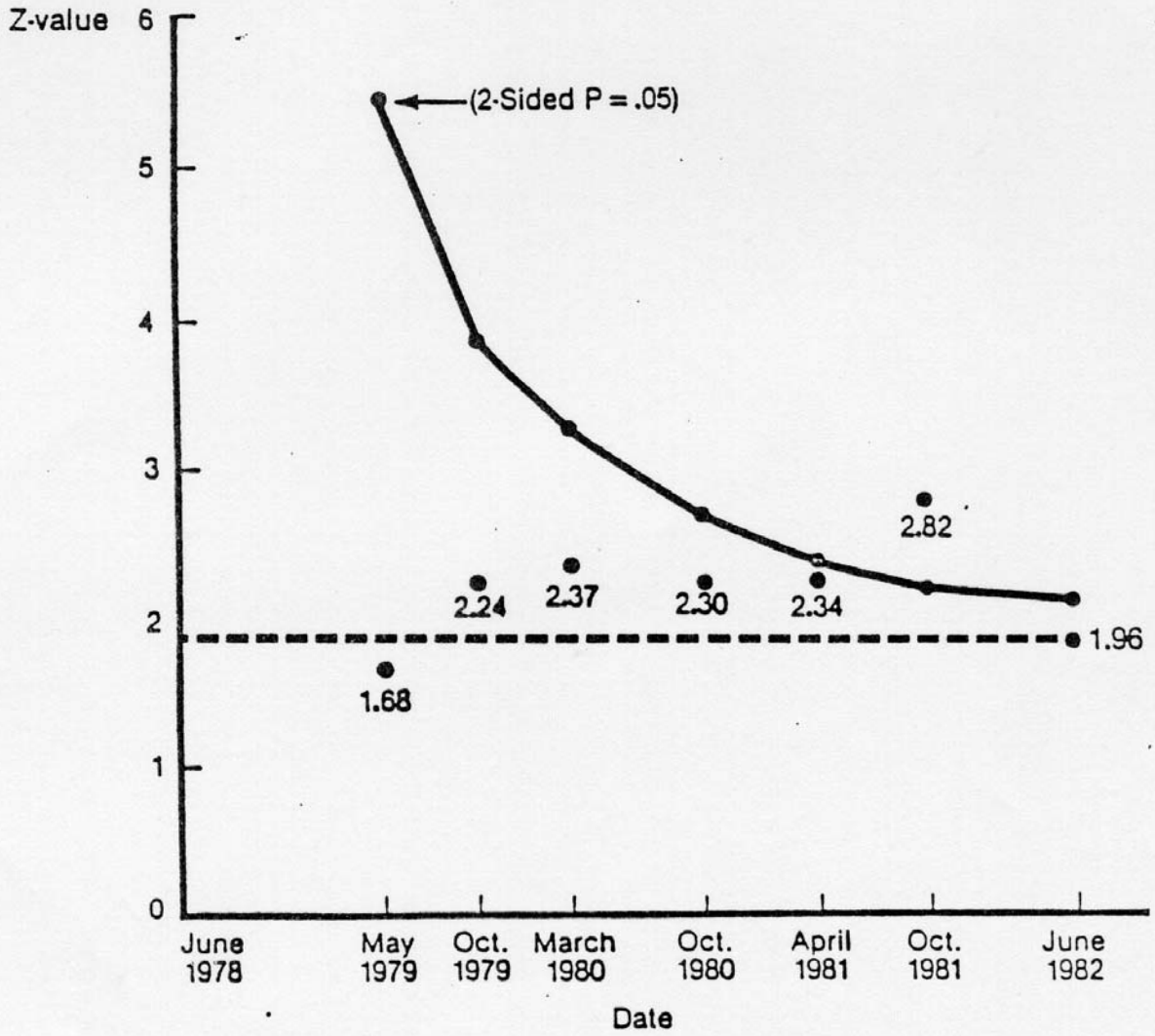
# LIFE-TABLE CUMULATIVE MORTALITY CURVES




log rank
Z = 2.82

Beta-Blocker Heart Attack Trial

# MONITORING BOUNDARY



# Cardiac Arrhythmia Suppression Trial (CAST) NEJM, 1989:321,406-412,1991: 24,781-788

- Design

  Dose titration enrichment phase plus parallel, placebo-controlled, randomized

- Treatment

  Flecainide, encainide, moricizine, and placebo

- Stratification

  LVEF (< 0.03 or >= 0.30), Holtering recording to MI, (<90 days or >= 90 days)

- Efficacy endpoints
  - Death or cardiac arrest from arrhythmia
  - Death or cardiac arrest from any cause

# Cardiac Arrhythmia Suppression Trial (CAST) NEJM, 1989:321,406-412,1991: 24,781-788

- Hypotheses

Does suppression of arrhythmia following an MI reduce incidence of

- Sudden death
- Total mortality

The spending function

$$\alpha^*(t) = \begin{cases} (\alpha^*/2)t, \text{ if } t < 1 \\ \alpha^*, \text{ if } t = 1 \end{cases}$$

t is the ratio of current number of events to the total number of events expected at the end of the trial

# Summary of the Interim Results Based on the Number of Deaths from Arrhythmia or Cardiac Arrest

| | Calendar Times for Interim Analyses | |
|---|---|---|
| | 9/1/88 | 3/30/89 |
| Placebo | 7(576)[a] | 9(725) |
| Active[b] | 22(571) | 33(730) |
| Total | 29(1147) | 42(1455) |
| Total Expected | 425 | 300 |
| One-sided $\alpha$ increment | 0.0009 | 0.0011 |
| Observed Logrank | -2.82 | -3.22 |
| Lower Bound[c] | -3.18 | -3.04 |

*Source:* Modified from Pawitan and Hallstrom (1990).

[a] Numbers in the parentheses are the number of patients assigned to the treatment.

[b] Active treatments include encainide and flecainide.

[c] Lower bounds were computed from a random sample with replacement of size b = 4000.

# U.S. Physician's Health Study
## NEJM 1989: 321 129-135

- Hypotheses
  - Does low dose aspirin (325 mg on alternate days) reduce cardiovascular
    - Mortality
    - Fatal and non-fatal infarction
  - Dose beta-carotene reduce incidence of cancer
- Design
  - Randomize 22000 male U.S. physicians
  - Factorial design
  - Double-blinded
  - Group sequential monitoring

    Statisticians: D. DeMets, T. Colton, L. Friedman

# U.S. Physician's Health Study

Five Years Results

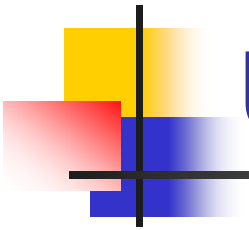
Mortality

| | Aspirin | Placebo | RR | P-value |
|---|---|---|---|---|
| Total | 217 | 227 | 0.96 | NS |
| CV | 81 | 83 | 0.96 | NS |
| MI | 10 | 28 | 0.31 | 0.004 |

# U.S. Physician's Health Study

Five Years Results

| | Aspirin | Placebo | RR | P-value |
|---|---|---|---|---|
| Nonfatal MI | 129 | 213 | 0.59 | <0.001 |
| Total Stroke | 119 | 98 | 1.22 | 0.15 |
| Ischemic | 91 | 82 | 1.11 | 0.50 |
| Hemorrhaged | 23 | 12 | 2.14 | 0.06 |



# U.S. Physician's Health Study

- Early Termination Issues
  - Low mortality
    - Target population ?
    - Clinical and statistical inference
  - Fatal + nonfatal MI
  - Stroke
  - Crossover following MI

# Decision to stop

- 15-month decision process
- Total mortality

  Observed 88 (44 vs. 44)

  Expected = 800
- Fatal & non-fatal MI

  RR = 0.53
- Moderate / stroke hemorrhaged stroke

  10 vs. 2
- Dropping after MI
- Decision is medical not statistical
  - Is MI a good surrogate for mortality

# Women Health Initiatives (WHI)

- US Physician Health Study: An all men trial
- Too few women in clinical trials
- Women Health Initiatives (WHI) by NIH director B. Healy (a cardiologist)in 1991
- Designed to run for 15 years to recruit 160,000 women with a price tag of US $ 727 millions by 2007
- One of the components is the WHI trial

# Women Health Initiatives (WHI)

- A randomizd, DB placebo-controlled primary prevention trial
- Investigate the benefits and risks of hormone replacement therapy (HRT)
- Estrogen + progestin vs. Placebo
- 16608 healthy postmenopausal women aged 50-79 years with intact uterus recruited between 1993 and 1998 with expectation of final analysis in 2005 after an average of approximately 8.5 years of follow-up

# Women Health Initiatives (WHI)

- A primary prevention trial
- Healthy subjects
- Low incidence, mortality and morbidity rates
- Large sample size and infeasible to repeat due to cost and long-term nature
- More comprehensive approach to monitoring the primary prevention trials
- Considerations of overall health benefit vs. risk into formal termination procedure

# Women Health Initiatives (WHI)

- Primary efficacy endpoint

  incidence of CHD
- Primary safety endpoint

  incidence of invasive breast cancer
- Competing benefits and risks

  colorectal cancer, hip fracture

  stroke, pulmonary embolism, endometrial cancer, death due to other causes

# Women Health Initiatives (WHI)

- A weight global (Freedman et al, 1996)

  $W = w_1 d_1 + \ldots + w_8 d_8$

  $d_i$: difference in proportions between the two groups for outcome i, i =1,…8

  $w_i$: weights for outcome i – expected proportion of diagnosed patients who will die of that disease within a specific years of diagnosis

Benefits and risks are not symmetric

# Women Health Initiatives (WHI)

A mixed approach to early termination

1. O-B boundaries for each of eight outcomes and for global index
2. Asymmetric upper and lower boundaries:

   one-sided $\alpha$ = 0.025 for benefit

   one-sided $\alpha$ = 0.05 for adverse effects

   adverse-effect boundaries were adjusted using Bonferroni method
3. Trial stops if the upper or lower boundaries were crossed and the result from global index was supportive at 0.20 level

# Women Health Initiatives (WHI)

Semiannual DSMB meeting since the fall of 1997

The tenth interim analysis on May 31, 2002

1. The weighted log-rank test statistic z = -3.19 crossed the lower boundary for adverse event z= -2.32
2. The global index is supportive (z = -1.62)
3. Additional evidence of risks on CHD, stroke, pulmonary embolism outweighed the evidence of benefit on hip fracture and colon cancer
4. DSMB recommended early termination of the estrogen plus progestin component of WHI

# Example of Ethical Dilemmas
## Controlled Clinical Trials, 1993: 14: 6-18

- The VA Cooperative Studies Program (CSP) Protocol 298
- Patient: AIDS free, HIV-infected
  Baseline: 200-500 cells/$mm^3$
- Design: Multicenter, double-blind, randomized, placebo-controlled
- Treatment: Zidovudine (AZT) at the time of randomization
  Placebo followed by AZT when CD4+ < 200 or AIDS develops

# The AIDS Clinical Trials Group (ACTG) - NIH

- Protocol 016
- Patient: Symptomatic HIV infected and
  CD4+: 200 - 800
- Design: Multicenter, double-blind,
  randomized, placebo-controlled
- Treatment: AZT (200 mg every 4 hours): 351
  Placebo: 360

# The AIDS Clinical Trials Group (ACTG) - NIH

- Protocol 019
- Patient: 1338 Asymptomatic HIV infected and
  CD4+: < 800
- Design: Multicenter, double-blind,
  randomized, placebo-controlled
- Treatment: AZT (100 mg 5 times daily)
  AZT (300 mg 5 times daily)
  Placebo: 360
- Stratification: CD+ < 200, CD+ 200-500, CD+ < 200

# Clinical Outcome Events for CSP 298 as of November 1989

| Outcome | Early Zidovudine (n = 149) | Late Zidovudine (n = 148) | P[a] |
|---|---|---|---|
| Death | 9[b] | 9 | 0.85 |
| AIDS | 12 | 18 | 0.34 |
| AIDS or death | 17 | 18 | 0.97 |

[a] P = Log-rank test results.

[b] Number of patients.

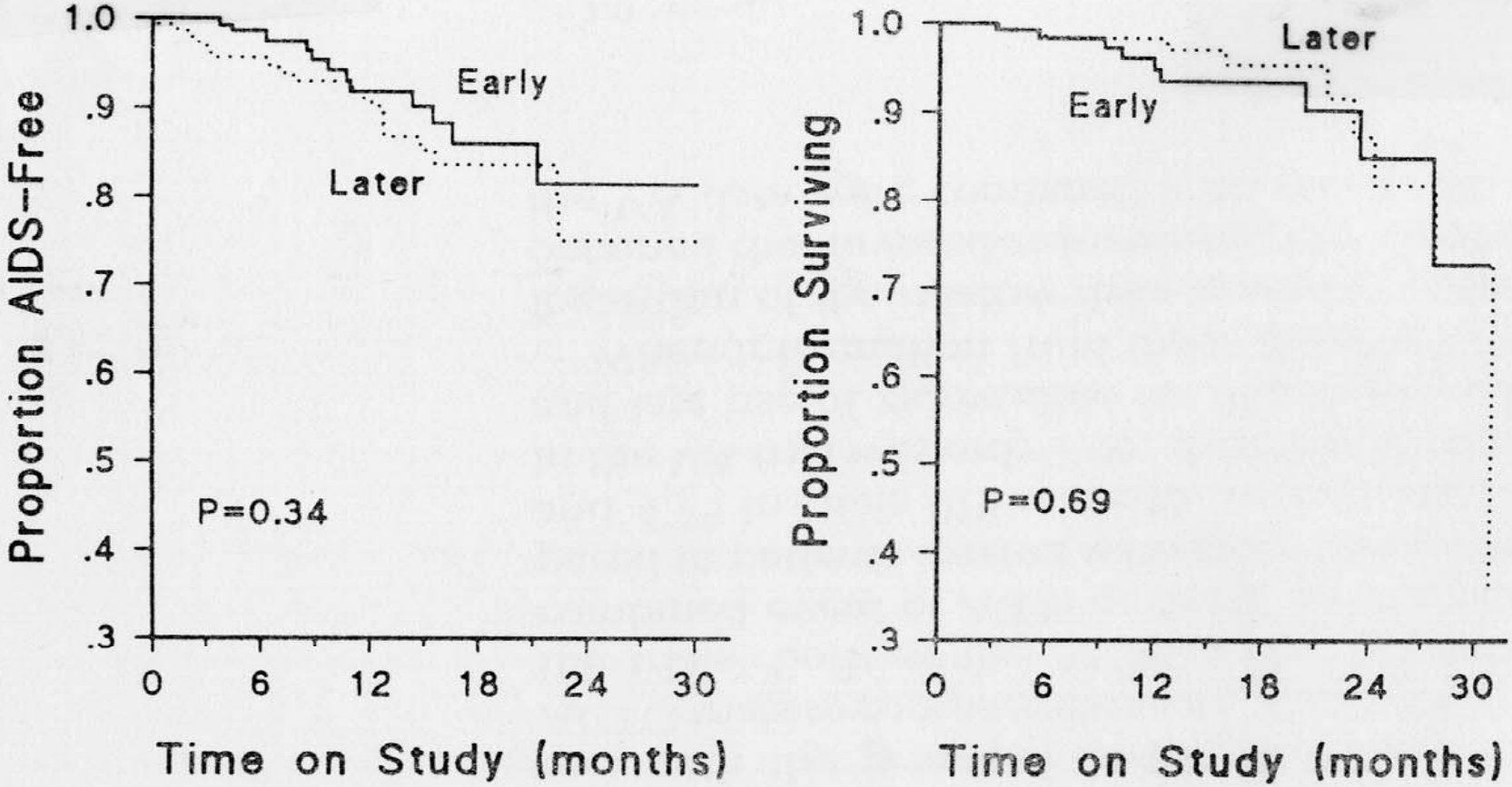



**Figure 1** Estimated Kaplan–Meier distribution of time to a diagnosis of AIDS (left) and time to death (right), according to trial group. Data are as of November 1989. Solid lines are patients randomized to early (immediate) zidovudine treatment and dashed lines are patients randomized to later (placebo-zidovudine) treatment.

**Table 2** Comparison of the ACTG and VA Studies for the Combined Outcome of Death or AIDS as of September 10, 1989 for Patients with Entry CD4+ Counts of 500

| Study | No. of Patients who Died or Developed AIDS | Rates/100 Person-Years of Follow-up | | Relative Risk (rr)[b] | $P$[c] |
|---|---|---|---|---|---|
| | | Placebo[a] | ZDV | | |
| ACTG Study 019 [6] | 58 | 6.6 | 2.7 | 2.45 | .0004 |
| ACTG Study 016 [5] | 23 | 8.2 | 1.7 | 4.75 | .002 |
| VA 298 [3] | 33 | 12.4 | 12.0 | 1.06 | .86 |
| (019 + 016 + 298) | 114 | | | 2.04 | .0003 |
| | | | | 95% CI (1.39, 3.01) | |
| (016 + 298) | 56 | | | 1.63 | .096 |
| | | | | 95% CI (.92, 2.90) | |

[a]"Placebo" for the ACTG studies but "later ZDV" for the VA Study.

[b]Relative risk is the rate ratio for 019 but the ratio of number of events ($D_1/D_2$) for 016 and 298.

[c]Significance levels are computed by treating

$$[\ln (rr)]^2 \left(\frac{1}{D_1} + \frac{1}{D_2}\right)^{-1}$$

as a $\chi^2$ variate with 1 d.f.

CI = Confidence interval; ZDV = zidovudine.

# Clinical Outcome Events in White and Nonwhite Patients for CSP 298[a]

| Outcome | White | | P[b] | Nonwhite | | P |
|---|---|---|---|---|---|---|
| | Early | Late | | Early | Late | |
| | No. of Patients | | | No. of Patients | | |
| Death | 4 | 9 | 0.47 | 5 | 0 | 0.048 |
| AIDS | 4 | 16 | 0.02 | 8 | 2 | 0.08 |
| AIDS or Death | 7 | 16 | 0.11 | 10 | 2 | 0.04 |

[a] Data as of November 1989.

[b] P = Log-rank test results.

The Global Utilization of Streptokinase and Tissue Plasminogen Activator for Occluded Coronary Arteries (GUSTO) – NEJM 1993: 329, 673-682

- Design
  - Four parallel groups
  - Randomized a total of 41021 patients
  - Multi-country and multicenter
  - Streptokinase (U.S. $ 320)
  - t-PA (U.S. $ 2300)
  - unblinded

# Group sequential monitoring of data

- A total ‘hand off’ independent DMC Statistician: D. DeMets, L. Fisher
- Pre-specified interim analysis of safety data for enrollment at 11,274,21,926, and 28,312 patients
- A two-sided symmetric O’Brien-Fleming boundaries generated with Lan-DeMets spending function.
- No allowance of conflict of interest in
  - Data monitoring committee
  - Steering committee
  - Data coordinating center

# Data Monitoring

- FDA guideline (1988, 2001)

"The process of examining and analyzing data accumulating ina a clinical trial, either formally or informally, can introduce bias. Therefore, all interim analyses, formal or informal, by any study participant, sponsor staff member, or data monitoring group should be described in full even if treatment groups were not identified. The need for statistical adjustment because of such analyses should be addressed. Minutes of meetings of the data monitoring group may be useful (and may be requested by the review division)."

# Standard Operating Procedures (SOP)

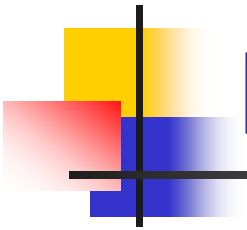
- Plans for interim analysis in trial protocol
- Methods for introducing un-planning interim analyses
- Methods for documenting interim analyses and their consequences on trial design and conduct
- Administrative cutoff of trials on information external to the trial
- Method for blinding and unblinding the trial

# General Issues

- Blinding and Un-blinding
  - Triple-blinded for confirmatory trials
  - Introduce bias
  - Establish SOP
    - Access to randomization code
    - When and how to unblind the data
    - Access to interim results
- Administrative Analyses
  - SOP for stopping trial early purely for administrative reasons
    - Who, Reasons, Actions and decisions
  - Potential Danger
    - Introduction of bias
    - Suggestion of trend but with no power for conclusive results
    - No adequate method available if the trial continues after administrative analyses

# Data monitoring – Review Interim Analyses

- Recruitment
- Baseline variables
- Clinical endpoints
  - Primary and secondary
- Toxicity / adverse events
- compliance



# Early vs. Late Analysis

- Early

  Mostly administrative

- Late

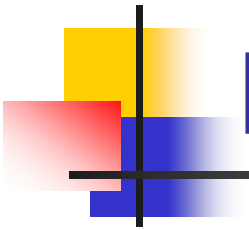
  Mostly early termination

# Interim Analysis in Early Stages of Drug Development

- Exploratory in nature
- For trial management
- Descriptive p-values
- Adequate documentation
  - Date of interim analyses
  - Number of patients
  - Summary statistics
  - Decisions

# Early Termination of Trials without adjusting p-value

- No potential early termination
- Major design flaws
  - Unusual placebo effect
  - Inappropriate outcome
- Unexpected toxicity
- Other independent reasons
  - Serious recruitment problems
  - Budget considerations
  - Administrative cutoffs
  - External information
    - Merge of companies
    - Trials done by CRO



# Early Administrative Analysis

- Recruitment / Entry Criteria
- Baseline comparison
- Design assumptions
  - Control only
  - Combined groups
- No possibility for an early termination

# Reasons for Early Termination

- Serious toxicity
- Established benefit
- No trend

# Interim Analysis in Late Stages of Drug Development

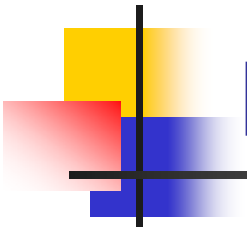
- Confirmatory trials of efficacy and safety for approval
- Few early termination for non-life-threatening condition
  - Clinical meaning not statistically significant differences
  - Aggregate and long-term safety
- Provide option of terminating trials early for ethical and scientific reasons

# Recommendations for Large Confirmatory Trials

- Group sequential procedures should be described in the protocol
  - Primary endpoints
  - Boundaries
  - References
  - Decision rules
- A Data Monitoring Committee should be established to review results of interim analyses
- Careful control of dissemination of results of interim analyses and deliberations of DMC to avoid bias
- Adequate documentation
  - Results of interim analyses
  - Minutes of DMC meetings

# Sample Size Re-estimation

- Treatment difference, variance, and control group response
- Reassessment of sample size
  - Overall response, overall variance
  - No adjustment of p-value at the end of study
  - Specify methods and decision in protocols
  - Perform at a scheduled interim analysis
  - Adequate documentation



# Data Monitoring Committee (DMC)

- NIH Model
- Disciplines
  - Clinical
  - Laboratory
  - Epideminology
  - Biostatistics
  - Data Managements
  - Ethics

# Data Monitoring Committee (DMC)

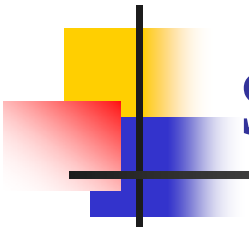
- Responsibility
  - Patients
  - Study / Investigators
  - Sponsor
  - Regulatory Agencies
- No conflict of interest

  No financial holding in the company
- Confidentiality

  No discussion of results outside DMC

# Statistical Analysis/Coordinating Center

- Independent of Sponsor
- Responsible for
  - Case Report Form design
  - Data entry
  - Quality control
    - Training
    - Certification
    - CRF tracking
    - Report
  - Interim analysis
  - Final analysis



# Statistical Analysis/Coordinating Center

- Interact closely with
  - Sponsor
  - Clinics
  - Steering Committee
  - Data monitoring Committee

# Decision Philosophy

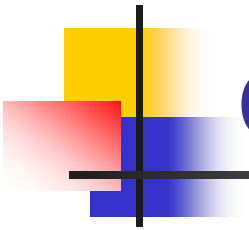
- Ahead of time
  - Positive trend
  - Negative trend
- Session Format
  - Open (blinded) Progress
    Recruiting, logistics
  - Closed (A vs. B) Clinical Endpoints
    Safety, Efficacy
  - Executive Decisions
    Who breaks the codes

# Decision Philosophy

- To whom does DMC report to
  - Sponsor
  - Study chair
  - Executive committee

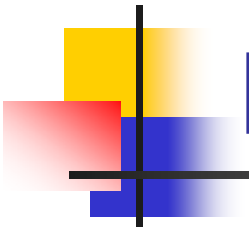
# DMC charges

- Protocol review
- Interim reviews
  - Study progress
  - Quality
  - Safety
  - Efficacy and Benefit
- Manuscript review – Primary results
  - Early termination



# On-line Data Management & Analysis

- DMC reluctant to make decision on old data
- Minimize data delay
  - Event verification
- Be prepare from start
- Focus on key endpoints not case report



# Documentation of DMC

- Protocol
  - Brief description
- Operations manual
  - Key endpoints
  - Decision process
  - Scenarios
- Interim data report / minutes

# DMC in Pharmaceutical Industry

- Pivotal phase Ⅲ trials
- Endpoints
  - Mortality
  - Irreversible morbidity
    - Myocardial infarction
    - stroke

# Methods Used

- Totally in-house
- External DMC
  - Internal analysis and process
- External DMC
  - Internal data process
  - External data analysis
- External DMC
  - External data process
  - External data analysis
- Totally “hand off”
  - Representative from sponsor on DMC

# External DMC

- Internal data process
- External data analysis
- Sponsor at open session only
- Sponsor at both open and closed session
  - Limited representation and confidentiality

# External DMC

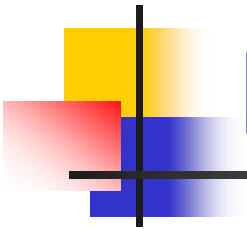
- Benefit
  - Industry manages large volume of data
  - Independent monitoring
  - Independent analysis
    - Verification of system
      - Case report form
      - Primary endpoint
- Total hand off
  - GUSTO trial

# Who attend DMC Meeting?

| Group | NIH | Industry |
|---|---|---|
| DMC | Yes | Yes |
| Study Chair | Yes | Yes |
| See study patients | No | No |
| Sponsor | Yes | Yes |
| Statistics | Yes | Yes |
| Regulatory agencies | No | No |

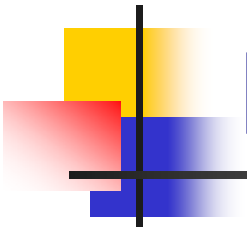
# Data Monitoring Committee

- Role of Regulatory Agencies
  - Should not participate
  - Could perhaps brief DMC
  - Should be briefed ASAP about DMC decisions
- Need for independent committee
  - Pivotal Phase Ⅲ studies
  - Resource not realistically available for all trials



# Discussion

- Single endpoint
  - Oversimplification of the study objectives
- Reproducibility of results at termination
- Bias in point and interval estimation
  - Estimation following sequential testing
- Designs with repeated measurements
- Design with more than two groups



# Discussion

- Sequential testing for equivalence trials
- Likelihood of reversal of results in a long-term study
- Difficult to communicate with clinicians adjusted p-values and penalty at the end if trials continue to their full extent
- Can different DMCs share confidential interim data (privacy vs. ethical)?

# References


- Chow, S.C. and Liu, J.P. (2005) Design and Analysis of Clinical Trials, 2nd Ed., Wiley, New York.
- Hybittle, J.L (1971) repeated assessment of results in clinical trials of cancer treatment, British Journal of Radiology, 44: 793-797
- Peto, R., Pike, M.C., Armitage, P., Breslow, N.E., Cox, D.R., Howard, D.R., Mantel, N., McPherson, K., Peto, J., Smith, P.G. (1976) Design and analysis of randomized clinical trials requiring prolonged observation of each patient. I. Introduction and design. British Journal of Cancer, 34: 585-612.
- Pocock, S.J. (1977) Group sequential methods in the design and analysis of clinical trials, Biometrika, 64: 191-199.

# References


- Fleming, T.R., Harrington, D.P., and O'Brien, P.C. (1984) Designs for group sequential tests, Controlled Clinical Trials, 5: 348-361.
- Jennision, C., and Turnball, B. (1991) Exact caculation for seequential t, chi-squre, and F tests, Biometrika, 78: 133-141.
- Jennision, C., and Turnball, B. (1989) interim analyses: The repeated confidence interval approach (with discussion), Journal of Royal Statistical Society, Series B, 51:305-361
- FDA (2001) Draft guidance on the establishing and operation of clinical trial data monitoring committee
- Freedman, L, et al (1998) Approaches to monitoring the results of long-term disease prevention trials: example from the Women Health Initiative, Controlled Clinical Trials, 17:509-529.
- O'Brien, P.C., and Fleming, T.R. (1979) A multiple testing procedure for clinical trials, Biometrics, 35: 549-556

# References


- DeMets, D.L., and Lan, K.K. G. (1994) Interim analysis: The alpha spending function approach, Statistics in Medicine, 13: 1341-1352.
- Lan, K.K.G., and Wittes, J. (1988) The B-value: A tool for monitoring data, Biometrics 44: 579-585.
- PMA Biostatistics and Medical Ad Hoc Committee on Interim Analysis (1993) Interim analysis in the pharmaceutical industry, Control Clinical Trials, 14: 160-173.
- Shih, W.J. (2001) Sample size re-estimation – a journal for a decade, Statistics in Medicine 20:515-518
- FDA (2001) Draft guidance on the Establishment and Operation of Clinical Data Monitoring Committee, Rockville, MD, USA.