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 observations 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 predetermined 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
  \[ s(t) = \frac{n(t)}{N(T_c)}, \quad 0 \leq t \leq T_c. \]
  $t_k$ is the calendar time at which the $k$th subject completes the study, $k = 0, 1, \ldots, N$
  \[ s_k = s(t_k) = \frac{n(t_k)}{N(T_c)}, \quad 0 \leq t \leq T_c. \]
  $s(t) = s_k$, if $t \in [t_k, \ t_{k+1})$
### Information Time vs. Calendar Time

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>6/1978</td>
<td>0 Months</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5/1979</td>
<td>11 Months</td>
<td>0.23</td>
<td>56</td>
<td>0.09</td>
</tr>
<tr>
<td>10/1979</td>
<td>16 Months</td>
<td>0.33</td>
<td>77</td>
<td>0.12</td>
</tr>
<tr>
<td>3/1980</td>
<td>21 Months</td>
<td>0.44</td>
<td>126</td>
<td>0.20</td>
</tr>
<tr>
<td>10/1980</td>
<td>28 Months</td>
<td>0.58</td>
<td>177</td>
<td>0.28</td>
</tr>
<tr>
<td>4/1981</td>
<td>34 Months</td>
<td>0.71</td>
<td>247</td>
<td>0.39</td>
</tr>
<tr>
<td>10/1981</td>
<td>40 Months</td>
<td>0.83</td>
<td>318</td>
<td>0.51</td>
</tr>
<tr>
<td>6/1982</td>
<td>48 Months</td>
<td>1.00</td>
<td>400/628</td>
<td>1.00</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>The Number of Repeated Test at the 5% level</th>
<th>Overall Significance level</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.05</td>
</tr>
<tr>
<td>2</td>
<td>0.08</td>
</tr>
<tr>
<td>3</td>
<td>0.11</td>
</tr>
<tr>
<td>4</td>
<td>0.13</td>
</tr>
<tr>
<td>5</td>
<td>0.14</td>
</tr>
<tr>
<td>10</td>
<td>0.19</td>
</tr>
<tr>
<td>20</td>
<td>0.25</td>
</tr>
<tr>
<td>50</td>
<td>0.32</td>
</tr>
<tr>
<td>100</td>
<td>0.37</td>
</tr>
<tr>
<td>1000</td>
<td>0.53</td>
</tr>
<tr>
<td>$\infty$</td>
<td>1.00</td>
</tr>
</tbody>
</table>

- 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

<table>
<thead>
<tr>
<th>Stage</th>
<th>Test Drug</th>
<th>Placebo</th>
<th>Information Fraction</th>
<th>Z-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>n/2</td>
<td>n/2</td>
<td>1/k(n)</td>
<td>Z1</td>
</tr>
<tr>
<td>2</td>
<td>2n/2</td>
<td>2n/2</td>
<td>2/k(2n)</td>
<td>Z2</td>
</tr>
<tr>
<td>3</td>
<td>3n/2</td>
<td>3n/2</td>
<td>3/k(3n)</td>
<td>Z3</td>
</tr>
<tr>
<td>K</td>
<td>kn/2</td>
<td>kn/2</td>
<td>1(kn)</td>
<td>Zk</td>
</tr>
</tbody>
</table>

At each of the K stages, compute the Z-statistics.
# Data Structure of Group Sequential Methods

<table>
<thead>
<tr>
<th>Number of Interim analysis (K)</th>
<th>Randomization</th>
<th>Information Time</th>
<th>Z- Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Test Drug</td>
<td>Placebo</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>n/2</td>
<td>n/2</td>
<td>1/k(n)</td>
</tr>
<tr>
<td>2</td>
<td>2n/2</td>
<td>2n/2</td>
<td>2/k(2n)</td>
</tr>
<tr>
<td>3</td>
<td>3n/2</td>
<td>3n/2</td>
<td>3/k(3n)</td>
</tr>
<tr>
<td>K</td>
<td>kn/2</td>
<td>kn/2</td>
<td>1(kn)</td>
</tr>
</tbody>
</table>

*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’Brien and Fleming (1979)

- 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. ± 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 ± 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 Turnbull (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
  \[ d_i \pm c_i \text{SE}(d_i) \]
- Offer something more than just p-value
The Conservative Critical Values

Two-sided: 0.05, One-sided: 0.025

<table>
<thead>
<tr>
<th>Number of Pocock</th>
<th>O’Brien-Fleming</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage (K)</td>
<td>Value</td>
</tr>
<tr>
<td>1</td>
<td>1.960</td>
</tr>
<tr>
<td>2</td>
<td>2.177</td>
</tr>
<tr>
<td>3</td>
<td>2.289</td>
</tr>
<tr>
<td>4</td>
<td>2.362</td>
</tr>
<tr>
<td>5</td>
<td>2.412</td>
</tr>
<tr>
<td>6</td>
<td>2.453</td>
</tr>
<tr>
<td>7</td>
<td>2.486</td>
</tr>
<tr>
<td>8</td>
<td>2.512</td>
</tr>
<tr>
<td>9</td>
<td>2.536</td>
</tr>
<tr>
<td>10</td>
<td>2.555</td>
</tr>
</tbody>
</table>

Modified from Jennison and Turnbull (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 an increasing function of information fraction $t$
  
  $\alpha^*(t) = \begin{cases} 
  0, & \text{if } t = 0 \\
  \alpha, & \text{if } t = 1 
  \end{cases}$

- 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

<table>
<thead>
<tr>
<th>$\alpha_1(s)$</th>
<th>Approximation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$2{1 - \varphi \left[ z\left( \frac{\alpha}{2}\right)/\sqrt{s} \right]}$</td>
<td>O'Brien-Fleming</td>
</tr>
<tr>
<td>$\alpha_2(s) = \alpha \ln[1+(e-1)s]$</td>
<td>Pocock</td>
</tr>
<tr>
<td>$\alpha_3(s) = \alpha s^\theta, \theta &gt; 0$</td>
<td>Lan-DeMets-Kim</td>
</tr>
<tr>
<td>$\alpha_4(s) = \alpha \left[ (1-e^{-\zeta s})/(1-e^{-\zeta}) \right], \zeta \neq 0$</td>
<td>Hwang-Shih</td>
</tr>
</tbody>
</table>
Examples of Boundaries by Alpha Spending Function

<table>
<thead>
<tr>
<th>Interim Analysis (s)</th>
<th>O’Brien-Fleming</th>
<th>$\alpha_1(s)$</th>
<th>pocock</th>
<th>$\alpha_2(s)$</th>
<th>$\alpha_3(s)[\theta -1]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1(0.2)</td>
<td>4.56</td>
<td>4.90</td>
<td>2.41</td>
<td>2.44</td>
<td>2.58</td>
</tr>
<tr>
<td>2(0.4)</td>
<td>3.23</td>
<td>3.35</td>
<td>2.41</td>
<td>2.43</td>
<td>2.49</td>
</tr>
<tr>
<td>3(0.6)</td>
<td>2.63</td>
<td>2.68</td>
<td>2.41</td>
<td>2.41</td>
<td>2.41</td>
</tr>
<tr>
<td>4(0.8)</td>
<td>2.28</td>
<td>2.29</td>
<td>2.41</td>
<td>2.40</td>
<td>2.34</td>
</tr>
<tr>
<td>5(1.0)</td>
<td>2.04</td>
<td>2.03</td>
<td>2.41</td>
<td>2.39</td>
<td>2.28</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Interim analysis(s)</th>
<th>Value</th>
<th>$\alpha$ (s)</th>
<th>Increment</th>
<th>Value</th>
<th>$\alpha$ (s)</th>
<th>Increment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1(0.2)</td>
<td>2.41</td>
<td>0.0079</td>
<td>0.0079</td>
<td>4.56</td>
<td>0.0000</td>
<td>2.6×10⁻⁶</td>
</tr>
<tr>
<td>2(0.4)</td>
<td>2.41</td>
<td>0.0138</td>
<td>0.0059</td>
<td>3.23</td>
<td>0.0006</td>
<td>0.000574</td>
</tr>
<tr>
<td>3(0.6)</td>
<td>2.41</td>
<td>0.0183</td>
<td>0.0045</td>
<td>2.63</td>
<td>0.0045</td>
<td>0.0039</td>
</tr>
<tr>
<td>4(0.8)</td>
<td>2.41</td>
<td>0.0219</td>
<td>0.0036</td>
<td>2.28</td>
<td>0.0125</td>
<td>0.0080</td>
</tr>
<tr>
<td>5(1.0)</td>
<td>2.41</td>
<td>0.0250</td>
<td>0.0031</td>
<td>2.04</td>
<td>0.0250</td>
<td>0.0125</td>
</tr>
</tbody>
</table>
Computation of Spending Probability and Boundaries

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

\[ P\{Z(0.2) > 4.56 \text{ or } Z(0.4) > 3.23\} \]
\[ = P\{Z(0.2) > 4.56\} + P\{Z(0.2) \leq 4.56 \text{ and } Z(0.4) > 3.23\} \]
\[ = 2.6 \times 10^{-6} + 0.000574 = 0.0006 \]

\[ P\{Z(0.2) > 4.56 \text{ or } Z(0.4) > 3.23 \text{ or } Z(0.6) > 2.63\} \]
\[ = P\{Z(0.2) > 4.56 \text{ or } Z(0.4) > 3.23\} + P\{Z(0.2) \leq 4.56 \text{ and } Z(0.4) \leq 3.23 \text{ and } Z(0.6) > 2.63\} \]
\[ = 0.0006 + 0.0039 = 0.0045 \]

\[ P\{Z(0.2) > 4.56 \text{ or } Z(0.4) > 3.23 \text{ or } Z(0.6) > 2.63 \text{ or } Z(0.8) > 2.28\} \]
\[ = P\{Z(0.2) > 4.56 \text{ or } Z(0.4) > 3.23 \text{ or } Z(0.6) > 2.63\} \text{ or } Z(0.8) > 2.28\} \]
\[ = 0.0045 + 0.0080 = 0.0125 \]

\[ P\{Z(0.2) > 4.56 \text{ or } Z(0.4) > 3.23 \text{ or } Z(0.6) > 2.63 \text{ or } Z(0.8) > 2.28 \text{ or } Z(1.0) > 2.04\} \]
\[ = P\{Z(0.2) > 4.56 \text{ or } Z(0.4) > 3.23 \text{ or } Z(0.6) > 2.63 \text{ or } Z(0.8) > 2.28\} \]
\[ + P\{Z(0.2) \leq 4.56 \text{ and } Z(0.4) \leq 3.23 \text{ and } Z(0.6) \leq 2.63 \text{ and } Z(0.8) \leq 2.28 \text{ and } Z(1.0) > 2.04\} \]
\[ = 0.0125 + 0.0125 = 0.0250 \]
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 $q = \frac{1.96 - B_t/t}{\sqrt{1-t}}$
- 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
    - 180 or 240 mg/day

## Beta-Blocker Heart Attack Trial

### Baseline Comparisons

<table>
<thead>
<tr>
<th></th>
<th>Propranolol (N=1,916)</th>
<th>Placebo (N=1,921)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Age (yrs.)</td>
<td>55.2</td>
<td>55.4</td>
</tr>
<tr>
<td>Male (%)</td>
<td>83.8</td>
<td>85.2</td>
</tr>
<tr>
<td>White</td>
<td>89.3</td>
<td>88.4</td>
</tr>
<tr>
<td>Systolic B.P.</td>
<td>112.3</td>
<td>111.7</td>
</tr>
<tr>
<td>Diastolic B.P.</td>
<td>72.6</td>
<td>72.3</td>
</tr>
<tr>
<td>Heart rate</td>
<td>76.2</td>
<td>75.7</td>
</tr>
<tr>
<td>Cholesterol</td>
<td>212.7</td>
<td>213.6</td>
</tr>
<tr>
<td>Current smoker (%)</td>
<td>57.3</td>
<td>56.8</td>
</tr>
</tbody>
</table>
## BHAT Accumulating Survival Data

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

* DMC terminated the BHAT
LIFE-TABLE CUMULATIVE MORTALITY CURVES

Cumulative Mortality Rate (%)

Months of Follow-up

Placebo

Propranolol

N = 3837  3696  3553  2850  2108  1202

log rank

\( z = 2.82 \)
Cardiac Arrhythmia Suppression Trial (CAST)

- **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)

Hypotheses
Does suppression of arrhythmia following an MI reduce incidence of
- Sudden death
- Total mortality

The spending function

\[ \alpha^*(t) = \begin{cases} 
\frac{\alpha^*}{2} t, & \text{if } t < 1 \\
\alpha^*, & \text{if } t = 1 
\end{cases} \]

\( t \) 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

<table>
<thead>
<tr>
<th>Calendar Times for Interim Analyses</th>
<th>9/1/88</th>
<th>3/30/89</th>
</tr>
</thead>
<tbody>
<tr>
<td>Placebo</td>
<td>7(576)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>9(725)</td>
</tr>
<tr>
<td>Active&lt;sup&gt;b&lt;/sup&gt;</td>
<td>22(571)</td>
<td>33(730)</td>
</tr>
<tr>
<td>Total</td>
<td>29(1147)</td>
<td>42(1455)</td>
</tr>
<tr>
<td>Total Expected</td>
<td>425</td>
<td>300</td>
</tr>
<tr>
<td>One-sided $\alpha$ increment</td>
<td>0.0009</td>
<td>0.0011</td>
</tr>
<tr>
<td>Observed Logrank</td>
<td>-2.82</td>
<td>-3.22</td>
</tr>
<tr>
<td>Lower Bound&lt;sup&gt;c&lt;/sup&gt;</td>
<td>-3.18</td>
<td>-3.04</td>
</tr>
</tbody>
</table>


<sup>a</sup> Numbers in the parentheses are the number of patients assigned to the treatment.

<sup>b</sup> Active treatments include encainide and flecainide.

<sup>c</sup> 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
## Five Years Results

### Mortality

<table>
<thead>
<tr>
<th></th>
<th>Aspirin</th>
<th>Placebo</th>
<th>RR</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>217</td>
<td>227</td>
<td>0.96</td>
<td>NS</td>
</tr>
<tr>
<td>CV</td>
<td>81</td>
<td>83</td>
<td>0.96</td>
<td>NS</td>
</tr>
<tr>
<td>MI</td>
<td>10</td>
<td>28</td>
<td>0.31</td>
<td>0.004</td>
</tr>
</tbody>
</table>
## Five Years Results

<table>
<thead>
<tr>
<th></th>
<th>Aspirin</th>
<th>Placebo</th>
<th>RR</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonfatal MI</td>
<td>129</td>
<td>213</td>
<td>0.59</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Total Stroke</td>
<td>119</td>
<td>98</td>
<td>1.22</td>
<td>0.15</td>
</tr>
<tr>
<td>Ischemic</td>
<td>91</td>
<td>82</td>
<td>1.11</td>
<td>0.50</td>
</tr>
<tr>
<td>Hemorrhaged</td>
<td>23</td>
<td>12</td>
<td>2.14</td>
<td>0.06</td>
</tr>
</tbody>
</table>
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 randomized, 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)

\[ W = w_1d_1 + \ldots + w_8d_8 \]

\( d_i \): difference in proportions between the two groups for outcome \( i \), \( i = 1, \ldots, 8 \)

\( w_i \): 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/mm³
- 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
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

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Early Zidovudine (n = 149)</th>
<th>Late Zidovudine (n = 148)</th>
<th>P&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Death</td>
<td>9&lt;sup&gt;b&lt;/sup&gt;</td>
<td>9</td>
<td>0.85</td>
</tr>
<tr>
<td>AIDS</td>
<td>12</td>
<td>18</td>
<td>0.34</td>
</tr>
<tr>
<td>AIDS or death</td>
<td>17</td>
<td>18</td>
<td>0.97</td>
</tr>
</tbody>
</table>

<sup>a</sup> P = Log-rank test results.

<sup>b</sup> 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

<table>
<thead>
<tr>
<th>Study</th>
<th>No. of Patients who Died or Developed AIDS</th>
<th>Rates/100 Person-Years of Follow-up</th>
<th>Relative Risk (rr)(^{a})</th>
<th>(p^{c})</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Placebo(^{a})</td>
<td>ZDV</td>
<td></td>
</tr>
<tr>
<td>ACTG Study 019 [6]</td>
<td>58</td>
<td>6.6</td>
<td>2.7</td>
<td>2.45</td>
</tr>
<tr>
<td>ACTG Study 016 [5]</td>
<td>23</td>
<td>8.2</td>
<td>1.7</td>
<td>4.75</td>
</tr>
<tr>
<td>VA 298 [3]</td>
<td>33</td>
<td>12.4</td>
<td>12.0</td>
<td>1.06</td>
</tr>
<tr>
<td>(019 + 016 + 298)</td>
<td>114</td>
<td>95% CI (1.39, 3.01)</td>
<td>95% CI (.92, 2.90)</td>
<td></td>
</tr>
<tr>
<td>(016 + 298)</td>
<td>56</td>
<td>2.04</td>
<td>.0003</td>
<td></td>
</tr>
</tbody>
</table>

\(^{a}\)"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\(^a\)

<table>
<thead>
<tr>
<th>Outcome</th>
<th>White No. of Patients</th>
<th>Nonwhite No. of Patients</th>
<th>(P)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Early</td>
<td>Late</td>
<td></td>
</tr>
<tr>
<td>Death</td>
<td>4</td>
<td>9</td>
<td>0.47</td>
</tr>
<tr>
<td>AIDS</td>
<td>4</td>
<td>16</td>
<td>0.02</td>
</tr>
<tr>
<td>AIDS or Death</td>
<td>7</td>
<td>16</td>
<td>0.11</td>
</tr>
</tbody>
</table>

\(a\) Data as of November 1989.

\(b\) \(P\) = Log-rank test results.

• 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 in 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
  - Epidemiology
  - 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?

<table>
<thead>
<tr>
<th>Group</th>
<th>NIH</th>
<th>Industry</th>
</tr>
</thead>
<tbody>
<tr>
<td>DMC</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Study Chair</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>See study patients</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Sponsor</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Statistics</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Regulatory agencies</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>
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


References

- FDA (2001) Draft guidance on the establishing and operation of clinical trial data monitoring committee
References

- PMA Biostatistics and Medical Ad Hoc Committee on Interim Analysis (1993) Interim analysis in the pharmaceutical industry, Control Clinical Trials, 14: 160-173.