Population Enrichment Designs
Case Study of a Large Multinational Trial

Harvard Schering-Plough Workshop
Boston, 29 May 2009

Cyrus R. Mehta, Ph.D
Cytel Corporation and Harvard School of Public Health

e-mail: mehta@cytel.com – web: www.cytel.com – tel: 617-661-2011
Acknowledgements

This research was performed jointly with:

- James Ware Ph.D. Harvard School of Public Health
- Ping Gao Ph.D. The Medicines Company
- Simona Skerjanec PharmD. The Medicines Company
- Deepak Bhatt MD, MPH. Brigham and Women's Hospital
- Robert Harrington MD. Duke Clinical Research Center

We thank Aniruddha Deshmukh for programming support, and Misha Salganik and Stuart Pocock for helpful discussions.
Motivation

- Cardiovascular disease is the most common cause of mortality and morbidity in the industrialized world
- But clinical trials of novel therapeutic agents face major challenges
  - low event rates
  - tiny effect sizes
  - diverse patient population

Such trials require enormous sample size commitments. Therefore it is advisable to build in the possibility for adaptive changes to the design based on interim looks.
Case Study: Platelet Inhibition during Percutaneous Coronary Intervention

- **Composite primary endpoint:** death, MI or ischemia driven revascularization within 48 hours

- **Control Arm:** Clopidogrel, a well studied ADP receptor antagonist binding to the P2Y$_{12}$ receptor

- **Experimental Arm:** Cangrelor, a novel ADP receptor antagonist expected to be more potent than Clopidogrel
Design Options Evaluated

1. Fixed Sample Design

2. Group Sequential Design

3. Adaptive Group Sequential Design with Sample Size Increase and Population Enrichment
1. Fixed Sample Design

- Sponsor decides on an initial commitment of \( N = 8000 \)
- This sample size provides more than 80% power if:
  - relative risk reduction \( \geq 18\% \) and control (clopidogrel) event rate \( \geq 10\% \)
- If these parameter estimates were be off by a few percentage points, study might be underpowered
Power Curves with Three Different Placebo Event Rates
Fixed Sample Designs (N = 8000)
Power Curves with Three Different Placebo Event Rates
Fixed Sample Designs (N = 8000)
Power Curves with Three Different Placebo Event Rates
Fixed Sample Designs (N = 8000)

- 8% Placebo Rate
- 10% Placebo Rate
- 8.7% Placebo Rate

Power

50% 55% 60% 65% 70% 75% 80% 85% 90% 95% 100%

Percentage Risk Reduction on the Treatment Arm

15.00% 16.00% 17.00% 18.00% 19.00% 20.00% 21.00% 22.00%
### Previous Trials of Clopidogrel

Table 1: Randomized Studies of Clopidogrel in Coronary Artery Disease

<table>
<thead>
<tr>
<th>Study</th>
<th>Primary Endpoint</th>
<th>Event Rate of Clopidogrel</th>
<th>Risk Reduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAPRIE</td>
<td>Death, MI, Stroke</td>
<td>5.3%</td>
<td>8.7%</td>
</tr>
<tr>
<td>CURE</td>
<td>Death, MI, Stroke</td>
<td>9.3%</td>
<td>20%</td>
</tr>
<tr>
<td>PCI-CURE</td>
<td>Death, MI, Urgent TVR</td>
<td>4.5%</td>
<td>30%</td>
</tr>
<tr>
<td>CREDO</td>
<td>Death, MI, Stroke</td>
<td>11.5%</td>
<td>26.9%</td>
</tr>
<tr>
<td>COMMIT</td>
<td>Death, Re-Infarction, Stroke</td>
<td>9.2%</td>
<td>9%</td>
</tr>
<tr>
<td>CHARISMA</td>
<td>Death MI, Stroke</td>
<td>6.8%</td>
<td>7%</td>
</tr>
</tbody>
</table>

From Meadows TA, Bhatt DL (2007) *Circulation Research,* 100, 1261-75
2. Group Sequential Design

- 8000 patients is a very large fixed commitment
- Can do better with a group sequential design
- Such a trial would terminate early if relative risk reduction is very large (efficacy) or very small (futility)
- But slightly larger up-front commitment (8750 subjects)
Group Sequential Design; 8750 Patients; Unconditional Type-1 Error is 0.025

Sample Size

Futility Zone

Efficacy Zone
Motivation for Adaptive Design

- Group sequential design is well powered if $RR \geq 20\%$
- But any $RR \geq 16\%$ is still clinically meaningful
- And at $RR = 16\%$, the design only has 63% power
- 14,000 subjects needed for 80% power at $RR = 16\%$
  - Sponsor is unwilling to place such a large bet for 16% RR
  - Trial highly overpowered in expected range $RR \geq 20\%$
- Can an adaptive strategy protect the power at $RR = 16\%$?
3. Adaptive Group Sequential Design with Population Enrichment

Event rates could be higher, and platelet inhibition of Cangrelor could be more potent in the higher-risk subgroups

- Plan to enroll 8750 patients as before
- Examine data after 6125 patients (70% information)
- If treatment benefit is confined to pre-specified subgroups
  - Restrict future enrollment to those sub-groups
  - Possibly increase the sample size for those sub-groups

- Such a data dependent change will pose statistical, logistical and regulatory challenges
Qualifications for Subgroups

- Likely to influence the event rates
- Experimental arm influenced more than control arm
- Easily identifiable in a clinical setting
- Consistently collected in the trial
Prior Selection of Subgroup

1. Using clinical judgement, create four non-overlapping partitions of the overall population

<table>
<thead>
<tr>
<th>Non-Overlapping Partition</th>
<th>% of Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>high-risk and clopidogrel-naive</td>
<td>30%</td>
</tr>
<tr>
<td>high-risk and pre-treated with clopidogrel</td>
<td>30%</td>
</tr>
<tr>
<td>low-risk and clopidogrel-naive</td>
<td>20%</td>
</tr>
<tr>
<td>low-risk and pre-treated</td>
<td>20%</td>
</tr>
</tbody>
</table>

high risk patients are those with diabetes and a troponin + marker
2. Form three nested subgroups from these four non-overlapping partitions:
   • $G_0$ = full population (100%)
   • $G_1$ = high-risk subgroup (60% of $G_0$)
   • $G_2$ = high-risk, clopidogrel-naive subgroup (50% of $G_1$)

3. The goal is to try and win with $G_0$; if not, then with $G_1$; if not, then with $G_2$
The Enrichment Strategy

Compute conditional power (CP) at interim analysis:

- If CP $\geq 80\%$, carry on with no change
- If CP $< 80\%$:
  - Try for 80% CP with $G_0$ patients and sample size increase (up to 16,000 patients)
  - If unable, try for 80% CP by enriching with only $G_1$ patients and sample size increase
  - If unable, try for 80% CP by enriching with only $G_2$ patients and sample size increase
- Terminate for futility if CP $< 20\%$ despite enrichment and sample size increase
Hypothesis Testing Strategy

• For populations \{G_0, G_1, G_2\}, let \{H_0, H_1, H_2\} be corresponding null hypotheses that treatment and control event rates are equal

• Consider three cases:
  – Case 1: population was not enriched at interim analysis
  – Case 2: population was enriched by \( G_1 \) at interim analysis
  – Case 3: population was enriched by \( G_2 \) at interim analysis

• A closed testing strategy will be employed so as to ensure strong control of type-1 error
Closed Testing Procedure at Final Analysis

Case 1
Did Not Enrich
  Adaptive Test of H0 (full population)

Case 2
Enriched with G1
  Test 1
  Adaptive Test of H0 ∩ H1 (full population)
  Test 2
  Standard Test of H1 (new population only)

Case 3
Enriched with G2
  Test 1
  Adaptive Test of H0 ∩ H2 (full population)
  Test 2
  Standard Test of H2 (new population only)
Graphical Depiction of Hierarchical Testing

- Test 1 is a level-$\alpha$ adaptive test of $H_0 \cap H_i$, performed on the combined data from Stage 1 ($G_0$ population) and Stage 2 ($G_i$ population)
  - If $G_i$ is also $G_0$ (i.e., no enrichment), no further testing
  - If Test 1 cannot reject, no further testing. It is a Gatekeeper!

- Test 2 is a level-$\alpha$ conventional test of $H_i$, performed only on the new data from Stage 2 ($G_i$ population)

- Both tests must reject in order to claim statistical significance
Performing an Adaptive Level-$\alpha$ Test

Suppose sample size is changed and population is enriched

- Compute the conditional type-1 error of the original (non-adaptive) design under the null hypothesis, say $\alpha^*$
- Make any desired change to the future course of the study, but preserve the $\alpha^*$
- Muller and Schafer (1999) have shown that this strategy will preserve the unconditional type-1 error of the trial
Muller and Schafer Principle

(1) The test statistic assumed the value -1.67 at look 1 and -1.88 at look 2

(2) The conditional probability under H0 that the test statistic will assume a value less than -2 at look 3 and thereby enter the efficacy zone is 0.22

(3) We may modify the design adaptively at the second look provided we preserve this conditional probability
Adaptive Sample Size Increase Keeping Conditional Type 1 Error Unchanged

Muller and Schafer Principle

(1) the sample size was increased at look 2 from 8750 to 10678 so as to achieve 80% conditional power

(2) Since the sample size increase was data dependant, the conditional type-1 error of 0.22 had to be preserved

(3) In order to preserve the conditional type-1 error at 0.22 the final critical value of the altered trial was raised from -2.00 to -1.93

Probability of entering efficacy zone is 0.22
Simulations: Risk Reduction is 16%

- Group Sequential Design: $N_{\text{max}} = 8750$, $E(N) = 7027$, Power = 63%

- Adaptive Enrichment Design:

<table>
<thead>
<tr>
<th>Risk Reduction</th>
<th>$G_0$</th>
<th>$G_1$</th>
<th>$G_2$</th>
<th>Power</th>
<th>Expected Sample Size</th>
<th>Type of Interaction</th>
<th>Win With Enrichment</th>
</tr>
</thead>
<tbody>
<tr>
<td>16%</td>
<td>21.1%</td>
<td>25.7%</td>
<td></td>
<td>78%</td>
<td>8502</td>
<td>Two Subgroups</td>
<td>17%</td>
</tr>
<tr>
<td></td>
<td>20.8%</td>
<td>20.8%</td>
<td></td>
<td>75%</td>
<td>8343</td>
<td>One Subgroup</td>
<td>13%</td>
</tr>
<tr>
<td></td>
<td>16%</td>
<td>16%</td>
<td></td>
<td>69%</td>
<td>7978</td>
<td>No Subgroup</td>
<td>5%</td>
</tr>
</tbody>
</table>

- Enrichment design boosts power by between 6% and 15% compared to conventional group sequential design
Simulations: Risk Reduction is 18%

- Group Sequential Design: \( N_{\text{max}} = 8750, \ E(N) = 6904, \ \text{Power} = 73\% \)

- Adaptive Enrichment Design:

<table>
<thead>
<tr>
<th>Risk Reduction</th>
<th>( G_0 )</th>
<th>( G_1 )</th>
<th>( G_2 )</th>
<th>Power</th>
<th>Expected Sample Size</th>
<th>Type of Interaction</th>
<th>Win With Enrichment</th>
</tr>
</thead>
<tbody>
<tr>
<td>18%</td>
<td>23.2%</td>
<td>27.6%</td>
<td></td>
<td>86%</td>
<td>8113</td>
<td>Two Subgroups</td>
<td>14%</td>
</tr>
<tr>
<td></td>
<td>23.4%</td>
<td>23.4%</td>
<td></td>
<td>84%</td>
<td>8000</td>
<td>One Subgroup</td>
<td>12%</td>
</tr>
<tr>
<td></td>
<td>18%</td>
<td>18%</td>
<td></td>
<td>79%</td>
<td>7683</td>
<td>No Subgroup</td>
<td>4%</td>
</tr>
</tbody>
</table>

- Enrichment design boosts power by between 6\% and 13\% compared to conventional group sequential design
Power Curves for Fixed-Sample and Group Sequential Designs

Percentage Risk Reduction

Power

FSD (N=8000)
GSD (Nmax=8750)
Power Curves for Fixed-Sample, Group Sequential, and Adaptive Design with Single Subgroup Interaction
Power Curves for Fixed-Sample, Group Sequential, and Two Adaptive Enrichment Designs

- FSD (N=8000)
- GSD (Nmax=8750)
- Adaptive GSD with Enrichment: Single Subgroup Interaction
- Adaptive GSD with Enrichment: Two Subgroup Interaction
Average Sample Sizes for Fixed Sample and Group Sequential Designs

- **FSD (N=8000)**
- **GSD (Nmax=8750)**

Percentage Risk Reduction

Average Sample Size
Average Sample Sizes for Fixed Sample, Group Sequential, and Adaptive Enrichment Designs with Single Subgroup Interaction.
Average Sample Sizes for Fixed Sample, Group Sequential, and Two Adaptive Enrichment Designs
Summary of Results

• If $RR \geq 20\%$ enrichment design is unnecessary as trial is adequately powered

• If $16\% \leq RR < 20\%$, enrichment design can recover power, provided subgroup interaction is present

• The greater the subgroup interaction, the more the enrichment design improves power

• Thus enrichment option acts as insurance against failure if treatment effect is small overall but large for some subgroups
Logistical and Regulatory Issues

• Create efficient, smoothly functioning processes of centralized randomization, electronic data capture and drug supply to recruit potentially 16,000 patients from 400 sites; handled by Duke Clinical Research Institute

• Form an independent interim analysis review committee (IARC) comprising clinicians and statisticians with relevant expertise

• Create a charter for the IARC with guidance for the adaptive changes, some flexibility to overrule if circumstances warrant, and mechanism for reporting adaptive decision
• Seek regulatory approval in good time prior to unblinding
  – Ask for a special protocol assessment (SPA)
  – For novel designs, ask for face to face meeting
  – Provide convincing evidence that design controls the type-1 error; theoretical arguments should be backed up by simulation results
  – Demonstrate that trial integrity will not be compromised
  – Ideally provide regulatory reviewers with a robust, friendly simulation tool for exploring operating characteristics of design
Concluding Remarks

• Why not run two independent trials; exploratory and confirmatory?
  – exploratory trial won’t be adequately powered for subgroups
  – if it fails, there won’t be any support for the confirmatory trial
  – current design contains an insurance policy for handling subgroups

• Pre-specification of subgroups not a statistical requirement. But important safeguard against chasing random noise

• Flexibility to change sample size and enrich population greatly increases complexity of trial management
  – Logistics of drug are complicated if sample size is increased
  – Recruitment might slow down if population is enriched

• Simulate the design thoroughly before committing to it


Proof that Test Procedure is Closed

\[ H_0 \cap H_1 \cap H_2 \]

\[ H_0 \cap H_1 \quad H_0 \cap H_2 \\ H_1 \cap H_2 \]

\[ H_0 \quad H_1 \quad H_2 \]

To reject any hypothesis you must also reject all intersections of that hypothesis with all other hypotheses
Case 1: If you don’t enrich, all these tests must be rejected in order to claim G0

If you don’t enrich, you must reject $H_0$ at level-alpha in order to win with G0. Test 1 (on the total population) is an adaptive level-alpha test of $H_0$. If it rejects, all the other tests indicated automatically reject also at level-alpha.
Case 2: If you enrich with G1, all these 
tests must be rejected to claim G1

If you enrich with G1 you must first reject \((H_0 \cap H_1)\) with Test 1, 
the “gatekeeper”. This is a level-alpha adaptive test. Then you must reject \(H_1\) 
with Test 2 using only the new G1 data. This is a conventional level-alpha test. 
If it rejects, all other tests indicated reject automatically at level-alpha
Case 3: If you enrich with G2, all these tests must be rejected to claim G2

If you enrich with G2 you must first reject \((H_0 \cap H_2)\) with Test 1, the “gatekeeper”. This is a level-alpha adaptive test. Then you must reject H2 with Test 2 using only the new G2 data. This is a conventional level-alpha test. If it rejects all other tests indicated automatically reject at level-alpha.
Verify by Simulation

Simulations for TheXYZ Company

Population

Parameters of Atomic Subgroups

<table>
<thead>
<tr>
<th>Parameter</th>
<th>HighR + Naïve</th>
<th>HighR + PreTrt</th>
<th>LowR + Naïve</th>
<th>LowR + PreTrt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subgr. Fr.</td>
<td>0.30</td>
<td>0.30</td>
<td>0.20</td>
<td>0.20</td>
</tr>
<tr>
<td>( \pi_c )</td>
<td>0.120</td>
<td>0.090</td>
<td>0.070</td>
<td>0.050</td>
</tr>
<tr>
<td>( \pi_l )</td>
<td>0.084</td>
<td>0.072</td>
<td>0.063</td>
<td>0.048</td>
</tr>
<tr>
<td>% Improvement</td>
<td>30.0%</td>
<td>20.0%</td>
<td>10.0%</td>
<td>4.0%</td>
</tr>
</tbody>
</table>

Subpopulations of interest

<table>
<thead>
<tr>
<th>Subpop.</th>
<th>Subpop. Fraction</th>
<th>Event Rates</th>
<th>% Improv.</th>
<th>Difference ( \delta )</th>
<th>Discount Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \pi_c )</td>
<td>( \pi_l )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( G_0 )</td>
<td>1.00</td>
<td>0.087</td>
<td>0.069</td>
<td>20.7%</td>
<td>0.018</td>
</tr>
<tr>
<td>( G_1 )</td>
<td>0.60</td>
<td>0.105</td>
<td>0.078</td>
<td>25.7%</td>
<td>0.027</td>
</tr>
<tr>
<td>( G_2 )</td>
<td>0.30</td>
<td>0.120</td>
<td>0.084</td>
<td>30.0%</td>
<td>0.036</td>
</tr>
</tbody>
</table>

\( G_0 \): All Patients, \( G_1 \): HighR Patients, \( G_2 \): HighR & Plavix Naïve Patients

Test

<table>
<thead>
<tr>
<th>Look #</th>
<th>SS (Both Arms)</th>
<th>Efficacy</th>
<th>Stopping Rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4000</td>
<td>Z ≥ 2.9626</td>
<td>% Impr. &lt; -1%</td>
</tr>
<tr>
<td>2</td>
<td>5600</td>
<td>Z ≥ 2.4623</td>
<td>CP &lt; 0.2</td>
</tr>
<tr>
<td>3</td>
<td>8000</td>
<td>Z ≥ 2.0018</td>
<td></td>
</tr>
</tbody>
</table>

Adaptation

Max. Sample Size With Adaptation (Both Arms; \( n_{max} \)) 15000

Adaptation Criterion: \( CP ≥ \) Upper Threshold 0.8

Approach to Sample Size Re-estimation

- Single Trial
- Separate Trials

Disable Enrichment

Total Number of Simulations 10000

Initial Random Number Seed clock
Power Curves for Fixed-Sample, Group Sequential and Adaptive Group Sequential Designs

- **FSD (N=8000)**
- **GSD (Nmax=8750)**
- **Adaptive GSD (N*=15,000)**

**Y-axis (Power):**
- 50%
- 55%
- 60%
- 65%
- 70%
- 75%
- 80%
- 85%
- 90%
- 95%

**X-axis (Percentage Risk Reduction):**
- 15.00%
- 16.00%
- 17.00%
- 18.00%
- 19.00%
- 20.00%
- 21.00%
- 22.00%
Power Curves for Fixed-Sample, Group Sequential, Group Sequential Adaptive and Enrichment under Homogeneous Treatment Effect

- FSD (N=8000)
- GSD (Nmax=8750)
- Adaptive GSD (N*=15,000)
- Adaptive GSD with Enrichment: Homogeneous Effect
Power Curves for Fixed-Sample, Group Sequential, Group Sequential Adaptive and Enrichment Design Under Homogeneous and Heterogeneous Treatment Effect

- FSD (N=8000)
- GSD (N_{max}=8750)
- Adaptive GSD (N^*=15,000)
- Adaptive GSD with Enrichment: Homogeneous Effect
- Adaptive GSD with Enrichment: Heterogeneous Effect

Percentage Risk Reduction vs. Power
Average Sample Sizes for Fixed Sample, Group Sequential Adaptive Group Sequential and Enrichment under Homogeneous Treatment Effect

Percentage Risk Reduction

Average Sample Size

- FSD (N=8000)
- GSD (N_{max}=8750)
- Adaptive GSD (N^*=15,000)
- Adaptive GSD with Enrichment: Heterogeneous Effect
Average Sample Sizes for Fixed Sample, Group Sequential, Adaptive Group Sequential, and Enrichment Designs under Homogeneous and Heterogeneous Treatment Effect

Percentage Risk Reduction

FSD (N=8000)
GSD (Nmax=8750)
Adaptive GSD (N* = 15,000)
Adaptive GSD with Enrichment: Homogeneous Effect
Adaptive GSD with Enrichment: Heterogeneous Effect
A Final Take-Away Message

Adaptive trials require a considerable amount of planning up-front. One of the most versatile tools for the planning phase is simulation:

- The simulations clarify the risks and benefits of the proposed approach by putting probabilities on each possible outcome.
- The simulations facilitate better communication with the FDA.
- The simulations facilitate greater communication between the various members of the study team by showing how different design options and trial outcomes will have different implications for:
  - patient recruitment
  - drug supply
  - economic analyses
  - clinical outcomes
  - statistical power
  - regulatory concerns