ADAPTIVE POPULATION ENRICHMENT DESIGNS IN CONFIRMATORY CLINICAL TRIALS

Vlad Dragalin, PhD
SVP Innovation Center
OUTLINE

1. Enrichment Strategies
2. Methodology
3. Examples
The aspect of "one size fits all" surrounding the conventional design of clinical trials has been challenged, particularly:

- when the disease is considered heterogeneous
- or the experimental therapy is tailored to a specific mechanism of action

One size fits all  Tailoring  Targeted Therapy

create diagnostic, prognostic and therapeutic strategies tailored for specific groups of patients
Patients Can Respond Differently

<table>
<thead>
<tr>
<th>Category</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>HYPERTENSION DRUGS</td>
<td>10-30%</td>
</tr>
<tr>
<td>ACE Inhibitors</td>
<td></td>
</tr>
<tr>
<td>HEART FAILURE DRUGS</td>
<td>15-25%</td>
</tr>
<tr>
<td>Beta Blockers</td>
<td></td>
</tr>
<tr>
<td>ANTI-DEPRESSANTS</td>
<td>20-50%</td>
</tr>
<tr>
<td>CHOLESTEROL DRUGS</td>
<td>30-70%</td>
</tr>
<tr>
<td>Statins</td>
<td></td>
</tr>
<tr>
<td>ASTHMA DRUGS</td>
<td>40-70%</td>
</tr>
<tr>
<td>Beta-2-agonists</td>
<td></td>
</tr>
</tbody>
</table>

Percentage of the patient population for which any particular drug is ineffective

The Case for Personalized Medicine
Edward Abrahams, Ph.D.¹ and Mike Silver, Ph.D.²
Journal of Diabetes Science and Technology
Volume 3, Issue 4, July 2009
The development of biomarkers that define specific subsets of disease is enabling a shift from empirical medicine to precision (stratified) medicine.

Trusheim, et. al, April 2007
A Paradigm Shift

Empirical Medicine
- Blockbuster drugs targeted at broad population segments
- On average, 50% of patients do not have desired therapeutic outcomes
- Significant adverse events

Precision Medicine
- Drugs targeted at subgroups of patient population
- Genomic profiles determine segmentation and therapy
- Best possible therapeutic outcome with minimal adverse events

Personalized Medicine
- Delivering the right medicine,
- to the right patient,
- at the right dose,
- at the right time

“Personalized Medicine means knowing what works, knowing why it works, knowing who it works for and applying the knowledge for patients” Michael Leavitt, Secretary of Health and Human Services

http://www.jyi.org/features/ft.php?id=1047
Potential Benefits

- Patients receive more effective drugs with fewer side effects giving better outcomes

- Avoid time and resources wasted trying unsuitable medicines

- Accelerating the development and availability of new diagnostics, medicines and treatment pathways benefit patients, healthcare providers and business.
Guidance for Industry

Enrichment Strategies for Clinical Trials to Support Approval of Human Drugs and Biological Products

DRAFT GUIDANCE

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)
Center for Devices and Radiological Health (CDRH)
December 2012
Clinical Medical
FDA Guidance: Adaptive Enrichment

V. PREDICTIVE ENRICHMENT

A. EMPIRIC STRATEGIES
   1. Open Trial Followed by Randomization
   2. An Individual’s History of Response to a Treatment Class
   3. Factors Identified in Results from Previous Studies

B. PATHOPHYSIOLOGICAL STRATEGIES
   1. Metabolism of the Test Drug
   2. Effect on Tumor Metabolism
   3. Proteomic Markers and Genetic Markers Linked to a Proteomic Marker

C. GENOMIC STRATEGIES

D. RANDOMIZED WITHDRAWAL STUDIES

E. STUDIES IN NON-RESPONDERS OR PATIENTS INTOLERANT TO OTHER THERAPY
   1. Studies in Non-Responders
   2. Study in Intolerants: Angiotensin Receptor Blockers (ARBs) in People Who Cough on Lisinopril

VI. ENRICHMENT STUDY DESIGN AND OTHER CONSIDERATIONS

A. GENERAL CONSIDERATIONS
   1. Performance Characteristics of a Screening Strategy for Selecting Patients
   2. When Should a Classifier Be Developed and Characterized?

B. WHICH POPULATIONS TO STUDY
   1. Studying Marker-Positive Patients Only
   2. Studying Both Marker Positive and Negative Patients
   4. Studies in Patients Intolerant of a Prior Treatment

C. TYPE I ERROR RATE CONTROL FOR ENRICHED STUDY SUBPOPULATIONS

D. ADAPTIVE ENRICHMENT

E. CAUTIONS IN INTERPRETATION
**Definition**

**Enrichment**

- is prospective use of any patient characteristic
  - demographic, pathophysiologic, historical, genetic, and others
- to select a study population in which
- detection of a drug effect (if one is in fact present) is more likely than it would be in an unselected population

*FDAC Draft Guidance for Industry: Enrichment Strategies for Clinical Trials, December 2012*
Reasons for enrichment

The main reason for enrichment is study efficiency

• increasing the chance of success, often with a smaller sample size
• providing major benefits of individualization,
• directing treatment where it will do the most good
• sparing potential harm to people who cannot respond
Key Concepts

- Extension from the conventional single population design objective to an objective that encompasses several possible patient sub-populations
- Allow more informative evaluation in the patients having different degrees of responsiveness to the therapy
- At an interim stage, it is decided which subpopulation is selected for further inference (including all subpopulations, i.e., full population)
- Not only selection procedures, but also other adaptive strategies (e.g., sample size reassessment, stopping rule) can be performed
EXAMPLE 1
I-SPY Model: A new paradigm in drug development

• Mainly focused on exploratory stage of DD
  – Match drugs with biomarker signatures
  – Savings from using a common control
  – Better therapies move through faster
  – Successful drug/biomarker pairs graduate to
    • small,
    • focused,
    • more successful Phase 3
  – based on Bayesian predictive probabilities
• Opens new opportunities in confirmatory stage of DD
A New Rx for Medicine

Fed up with slow drug trials, cancer patients and doctors are testing a fast track to personalized treatments.

By RON WINSWOT

PERSONALIZED MEDICINE | How redesigning a clinical trial can speed drug development

Traditional clinical trial
Takes essentially all patients with a disease being studied and is typically intended to eliminate differences in patient characteristics that could bias measures of drug effectiveness.

New trial design
Uses genetic profiles to highlight ‘biomarker’ differences among patients and to match drugs to patients with biomarkers that predict a benefit.

PHASE II
Randomized or non-randomized trial: In a randomized trial, about 60 patients are put in two groups: One receives the experimental drug and the other serves as a control group. In a non-randomized trial, about 40 patients receive the experimental drug.

PHASE III
If a drug graduates to phase III, it typically takes 3,000 patients and about three years to determine if it is safe and effective enough for approval.

HISTORIC SUCCESS RATE
30% to 40%

PHASE II
Patients are placed in groups based on genetic profiles and are randomly assigned to either standard therapy or one of five different drugs plus standard care.

Early results increase chances that patients entering the trial later will be assigned to a drug showing benefit against tumors with their genetic profile.

It will take up to 120 patients for each drug to determine which ones graduate to phase III studies.

PHASE III
Researchers expect that drugs graduating from Phase II to phase III can be tested with 300 patients selected according to genetic profiles found to respond to the drug in phase II. It is hoped that this will shorten the time to approval.

PROBABILITY OF SUCCESS
85%

Source: Donald Berry, M.D., Anderson Cancer Center
Phase 3 Study in HER2- MBC Patients

- Assume that one of the experimental drugs has been graduated from the I-SPY 2 trial with the biomarker signature of triple negative breast cancer (TNBC) but also with some promising effect in HER2- biomarker signature.

- **Option 1**: a confirmatory Phase 3 trial in TNBC patients only
  - prevalence of TNBC is only 34%

- **Option 2**: a confirmatory Phase 3 trial in HER2- patients
  - prevalence of HER2- is 63%

- **Option 3**: Adaptive enrichment design
  - run a confirmatory trial with a two-stage enrichment design
  - starting with the full population (HER2- patients),
  - but with the preplanned option of selecting only the TNBC patients after the 1st stage in case the observed effect is not promising in the HER2- patients with positive hormone-receptor status HR+

### Acknowledgment
D. Berry. I-SPY-1 Results
- **Stage 1 objective**
  - Stop for futility/efficacy
  - To continue with HER2- (Full) population
  - To confirm greater benefit in TNBC Subpopulation (Sub)
  - To adjust the sample size

- **Stage 2 data and the relevant groups from Stage 1 data combined**
Ballpark Sample Size Calculations

- **Primary Endpoint**: pathologic complete response (pCR) at surgery
- **Power**: 90%
- **Sign. Level**: 0.025
- **Control Rate**: pCR=0.3
- **TRT Effect**: 0.2

Possible TRT Effect Range: [0.1 – 0.25]
Population Enrichment Simulation

<table>
<thead>
<tr>
<th>Patient Profile</th>
<th>MP- Her2+ HR+</th>
<th>MP- Her2+ HR-</th>
<th>MP- Her2- HR+</th>
<th>MP- Her2- HR-</th>
<th>MP+ Her2+ HR+</th>
<th>MP+ Her2+ HR-</th>
<th>MP+ Her2- HR+</th>
<th>MP+ Her2- HR-</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prevalence</td>
<td>16%</td>
<td>7%</td>
<td>23%</td>
<td>6%</td>
<td>4%</td>
<td>10%</td>
<td>6%</td>
<td>28%</td>
</tr>
<tr>
<td>Predicted pCR</td>
<td>47%</td>
<td>67%</td>
<td>25%</td>
<td>43%</td>
<td>35%</td>
<td>55%</td>
<td>17%</td>
<td>32%</td>
</tr>
</tbody>
</table>

Acknowledgment: D. Berry. I-SPY-1 Results

- Prevalence of TNBC in HER2- : 54%
- Control pCR Rate in TNBC: 0.34
- Control pCR Rate in HER2- ∩ HR+: 0.23
- Total of 21 Simulation Scenarios:
  - TRT effect in TNBC: 0 to 0.3 by 0.05
  - TRT effect in HER2- ∩ HR+: 0, 0.1, 0.2

Design
- Total sample size: 300 patients
- Stage 1 sample size: 150 pats
- Testing strategy: inverse normal p-value combination
- Intersection test: Bonferroni
- Selection rule: $\varepsilon = 0.1$ rule
Operating Characteristics:

Power | P_Reject F | P_Reject S1 vs. Effect S1

Pitch Subset2 = 0.230
Operating Characteristics:

Legend
- **Power**
- P_Reject F
- P_Reject S1

Power | P_Reject F | P_Reject S1 vs. Effect S1

\[\pi T_{\text{Subset2}} = 0.330\]
Operating Characteristics:

Legend
- Blue: Power
- Orange: P_Reject F
- Red: P_Reject S1

Power | P_Reject F | P_Reject S1 vs. Effect S1

piT Subset 2 = 0.430
Sample Size Reestimation

- Allow up to a 3-fold sample size increase for Stage 2
- 90% Conditional Power based on observed TRT effect
- Total Sample Size: 300 - 600
Operating Characteristics

Power | P_Reject F | P_Reject S1 vs. Effect S1

piT Subset2 = 0.430

Legend
- Blue: Power
- Orange: P_Reject F
- Red: P_Reject S1
METHODOLOGY
Adaptive Confirmatory Designs

All information available in an interim analysis may be used for planning the subsequent stages of the trial, under control of the prespecified Type I error rate.

Two pioneering proposals:

1. Bauer & Köhne (Biometrics, 1994):
   Combination of $p$-values with a specific combination function (Bauer, 1989)

2. Proschan & Hunsberger (Biometrics, 1995):
   Specification of a conditional error function
Procedure of Bauer & Köhne (1994)

Stage 1:
- $p_1$
- $0 \to \alpha_1 \to \alpha_0 \to 1$
- Rejection of $H_0$ at Stage 1
- Acceptance of $H_0$

Stage 2:
- $p_1 p_2$
- $0 \to c_\alpha \to 1$
- Rejection of $H_0$ at Stage 2
- Acceptance of $H_0$
Procedure of Bauer & Köhne (1994)

- Use of Fisher’s combination test to combine the separate stage $p$-values $p_1$ and $p_2$, i.e., $C(p_1, p_2) = p_1 p_2$

- Under $H_0$, the $p$-values are stochastically independent, irrespective of the choice of the design for the second stage.

- $H_0$ is rejected after the second stage if

$$p_1 p_2 \leq c_\alpha = \exp(-1/2 \chi^2_{4,\alpha})$$

- Other combination functions $C(p_1, p_2)$ and/or more than two stages can also be considered.

- In the two stages, different hypotheses can be considered, the considered global test is a test for $H_0 = H^1_0 \cap H^2_0$
Adaptive Design using the *inverse normal method*

Consider at $k$th stage, $k = 1,2,\ldots,K$:

$$T_k^* = C(p_1,\ldots,p_k) = \frac{\Phi^{-1}(1 - p_1) + \Phi^{-1}(1 - p_2) + \ldots + \Phi^{-1}(1 - p_k)}{\sqrt{k}}$$

$\Phi^{-1}(1 - p_k) \sim N(0;1)$ if $p_k$ uniformly distributed on $[0; 1]$

Under $H_0$, the same distributional assumption as for the group sequential tests applies and, therefore, the decision regions of the traditional group sequential tests can be used.

Lehmacher & Wassmer, 1999
Properties

- Decision regions of group sequential tests can be used
- Generalization to more than two stages and more general designs straightforward
- Use *unweighted* mean of test statistics from the separate stages also for unequal and arbitrarily (data dependent) fixed sample sizes.
- Effect on power is small unless “dramatic“ changes in sample size were performed
- Can also be used in testing situations with nuisance parameters
- If no design changes were performed, the inverse normal technique yields the traditional test
Methodology for Population Enrichment

• Sources for alpha inflation
  – Interim analyses
  – Sample size reassessment
  – Selection from multiple sub-populations

• The adaptive procedure strongly controls the pre-specified family-wise Type I error rate

• The procedure is based on the application of the closed test procedure together with combination tests
Consider prespecified subpopulation(s) $S_1, \ldots, S_G$, which can be nested, and a full population $F$:

$$S_G \subset \ldots \subset S_1 \subset F$$

The proposed adaptive procedure fulfills the regulatory requirements for the analysis of adaptive trials in that it strongly controls the prespecified (familywise) Type I error rate.
Closed testing procedure

Stage I

\[ \begin{align*}
H_0^F \cap H_0^{S_1} \cap H_0^{S_2} \\
H_0^F \cap H_0^{S_1} \quad H_0^F \cap H_0^{S_2} \\
H_0^{S_1} \quad H_0^{S_2}
\end{align*} \]

Stage II

\[ \begin{align*}
? \quad ? \quad ? \\
H_0^S
\end{align*} \]

Simple “trick”: Test of intersection hypotheses are formally performed as tests for \( H_0^S \).

\( H_0^S \) can be rejected if all combination tests exceed the critical value \( u_2 \).
Closed testing procedure: Stage II

Example $S = S_2$

Stage I

$H_0^F \cap H_0^{S_1} \cap H_0^{S_2}

H_0^F \cap H_0^{S_1}

H_0^F

H_0^{S_1}

H_0^{S_2}

Stage II

$H_0^{S_2}$ can be rejected if all combination tests exceed the critical value $u_2$.

The choice of tests for intersection hypotheses is free. You might use Bonferroni, Simes or Sidak tests.

For one subgroup also Dunnett’s test can be applied
Test strategies

- **Combination test:**
  - Inverse normal method
  - Fisher’s combination test

- **Separate Phase II/III:**
  - Phase II only for sub-population selection
  - Phase III is group sequential

- **Intersection Tests:**
  - Dunnett
  - Bonferroni
  - Sidak
  - Simes
  - Hierarchical
Selection Procedure

- Select the (sub)population with the largest effect
- Select $r$ sets with largest effect
- Select sets with effect compared to full population not worse than $\varepsilon$
- Select $i$-th set
- Select a set if effect exceeds a threshold $t$
- Drop a set if $CP < x$
- Effect measured on test statistic or mean effect scale
Different Configurations

- S1: 64%
- S2: 20% (8% overlap)
- S2: 40%
- S1: 80%
- S1: 25% (20% overlap)
- S1: 20%
- S2: 20% (8% overlap)
- S3: 20%
EXAMPLE 2
Simulation Study: Continuous Endpoint

- Two-stage design with no early stopping, one sub-population
- In the biomarker positive population a standardized effect of 0.5 is assumed, biomarker negative population has effect sizes ranging from 0 to 0.5
- Selection rules:
  1. Select the population with highest effect size
  2. Select the population with effect size compared to the better not worse than 0.25
  3. No selection

- Prevalances of biomarker positive population is 5%, 10%, 20%.
- Sample size: 100 patients per stage
- Simes' test is used for testing intersection hypotheses.
ADDPLAN Base is a comprehensive package for the planning, simulation and analysis of clinical trials, comprising

- traditional fixed sample designs
- group sequential designs: O'Brien & Fleming, Pocock, Wang & Tsiatis, Pampallona & Tsiatis
- \( \alpha \)- and \( \beta \)-spending function approaches: Kim & DeMets, Pocock, O’Brien & Fleming, Hwang, Shih & DeCani, user defined \( \alpha \)-spending and \( \beta \)-spending
- Flexible (adaptive group sequential) designs:
  - inverse normal designs: Lehmacher & Wassmer, Cui, Hung & Wang,
  - p-value combination designs: Bauer&Koehne
  - conditional error function designs: Proschan&Hunsberger, Muller&Schafer
Specifications: Stopping Rule

- Number of stages: K = 2
- Group Sequential Design
  - Stopping for futility
  - Choice of design:
    - Pocock's design (\(\Delta=0.5\))
    - O'Brien and Fleming's design (\(\Delta=0\))
    - Choose Delta
    - Optimum Delta
    - Pampallona and Tsiatis design
    - Specify alpha spending
    - Specify alpha and beta spending
- Information rates:
  - Stage 1: 2
  - Rates: 0.5, 1.0
  - No interim stops
Specifications: Selection Rule

- Selection procedure
  - Select set (incl. full population) with largest effect
  - Select the r sets with largest effect, r = 2
  - Select sets with effect compared to best not worse than epsilon = 0.25
  - Select sets with effect compared to full not worse than epsilon
  - Select the ith set (incl. full population i = G), i = 1
  - Deselect sets (incl. full population) for which effect smaller epsilon
  - p-q-selection rule
    - p = 1.0
    - q = 1.0

- Effect measure
  - treatment difference
  - test statistic

- Stopping for success criterion
  - if effect is shown in all selected analysis sets
  - if effect is shown in at least one selected analysis sets

- Threshold condition
  - Select analysis set unconditionally
  - Select analysis set if effect exceeds the threshold t
### Specifications: Sample Size Reestimation

#### Sample Size Specifications

Preplanned overall sample size per stage

<table>
<thead>
<tr>
<th>Stage</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

Stage 1 sample size allocation $n_T/n_C = 1.0$

#### Sample Size Recalculation

- **No sample size recalculation**

- **Sample size recalculation with conditional power**
  - Maximum relative reduction $n$ per stage = 0.5
  - Maximum relative increase $n$ per stage = 4
  - Conditional power for next stage = 80%
  - Overall conditional power = 80%

Conditional power calculation based on

- Observed effect (ML estimate)
- Assumed standardized effect =
Specifications: Scenarios for Simulation

- **# of analysis sets**
  - G = 2

- **Effective analysis set**
  - Treatment effective if effect > 0

- **Effect specification**
  - **Prevalences**
    - S1: 10%
    - Rem: 90%
  - **Effect sizes**
    - from 0.5 to 0.5 by 0.1
    - from 0 to 0.5 by 0.1
  - **std** = 1

- **Free combination**
- **Specify effect separately**
  - All subsets
  - Subset S1
  - Remaining
Power plot

Power vs. Effect subset 2
Prevalences = (20 80)

Selection rule:
- Select set with largest effect at stage 1 (difference)
- Select the 2 sets with largest effect at stage 1 (difference)
- Select sets with effect compared to best not worse than epsilon at stage 1 (difference)
Power plot

Power vs. Effect subset 2
Prevalences = (10 90)

Selection rule
- Select set with largest effect at stage 1 (difference)
- Select the 2 sets with largest effect at stage 1 (difference)
- Select sets with effect compared to best not worse than epsilon at stage 1 (difference)
Power plot

Power vs. Effect subset 2
Prevalences = (5 95)

Selection rule
- Select set with largest effect at stage 1 (difference)
- Select the 2 sets with largest effect at stage 1 (difference)
- Select sets with effect compared to best not worse than epsilon at stage 1 (difference)
Results

• Clear power disadvantage for procedure that never selects a sub-population
• No clear advantage of selecting always (and only) the best population
• For small prevalences, always selecting the best can even provide a small loss in power
Summary

- Attractive and general procedure for adaptive confirmatory design that controls Type I error rate
- The “rules” for adaptation and stopping for futility
  - Do not need to be pre-specified
  - Adaptations may depend on all interim data including secondary and safety endpoints.
  - Can make use of Bayesian principles integrating all information available, also external to the study
  - Should be evaluated (e.g. via simulations) and preferred version recommended, e.g., in the Simulation Report or DMC Charter
- Comparison of different strategies and options for analyses is mandatory. The role of simulation becomes increasingly important
QUESTIONS ???