Enrichment Design with Patient Population Augmentation

Bo Yang, AbbVie
Yijie Zhou, AbbVie
Lanju Zhang, AbbVie
Lu Cui, AbbVie

2015 Duke-Industry Statistical Symposium
**Author Disclosure**

- Bo Yang is a former Employee of AbbVie, Inc. All other authors are employees of AbbVie, Inc.

- The presentation was sponsored by AbbVie. AbbVie contributed to the developing, writing, reviewing and approving the publication. The presentation contains no AbbVie data.
Abstract

The advancement in science (e.g., biomarkers from genomics, proteomics) has provided opportunities to identify patient sub-populations that may be more responsive to a treatment. Clinical trials can be enriched on such sub-populations to improve the probability of success in demonstrating the benefit of new treatments. In 2012 FDA issued a draft guidance document to facilitate trial design with an enrichment strategy. In this talk, we consider an enrichment trial design where efficacy in the enriched population and further in general patient population can be evaluated. Specifically, a new weighted test statistic is derived to assess the treatment effect in a general patient population under an enriched trial setting, coupled with novel design based on screening information for weight determination. The proposed design and analysis method enhance the probability of success compared with a traditional all-comer trial design. It allows a generalization of the enrichment trial result to an all-comer population despite the fact the enriched population is disproportionally distributed in the study. Sample size determination and data collection are discussed along with examples and simulations.
Outline

• Introduction
• Review of Enrichment Design
• Biomarker-positive Enrichment Strategy
• Biomarker-stratified Enrichment Strategy
• Enrichment design with patient population augmentation
• Simulation Result
• Conclusion
Introduction

• Traditional All-Comer trial focuses on the average treatment effect in overall population

• Many new agents, especially in cancer area are molecularly targeted
  • Might only benefit a subgroup of patients
  • All-Comer trial are suboptimal for evaluation of these targeted treatments

• Enrichment design utilizing biomarkers to identify beneficial subpopulation can both improve patient care and accelerate drug development
Enrichment Strategy

- 2012 FDA guidance: Enrichment Strategies for Clinical Trials to Support Approval of Human Drugs and Biological Products
  - Predictive enrichment strategy by targeting marker+ patients
- Successful enrichment improves the efficiency of a trial design
  - Increase the power of the study
  - Minimize the required sample size or duration
  - Enhance study POS
Consideration in Enrichment Strategy

- Potential increase in efficiency with enrichment design should be considered against the reduction in generalizability of the results

- Enrichment design targeting the subgroup of biomarker-positive (biomarker+) patients depends on the confidence in Biomarker's performance

- Often times, the development of targeted agents and their accompanying biomarkers occurs simultaneously
Biomarker+ Enrichment Strategy

- Study of biomarker+ subpopulation only
- Appropriate design if treatment benefit is likely to be in biomarker-positive subgroup only
- Trastuzumab development for adjuvant use in breast cancer
  - Two successful enrichment studies targeting 20% HER2-positive patients
Biomarker+ Enrichment Strategy: Issues

- Possible treatment effect in the biomarker-negative (biomarker-) subgroup
  - Multiple pathways related to treatment effect
  - Accuracy of biomarker assay in identify patients expressing the target
  - Uncertainty around cutoff with biomarker

- Cost adjustment and inconvenience in identifying the biomarker+ patients for treatment in routine clinical practice
Biomarker+ Enrichment Strategy

• Need for Inclusion of appropriate number of biomarker-patients in such enrichment strategy

• Trastuzumab development for adjuvant use in breast cancer
  • Evidence of treatment benefit in patients historically classified as HER2-negative despite having some expression of HER2
  • Ongoing 3260 patients P III study to assesses the benefit of adjuvant trastuzumab in invasive breast cancer patients with positive, but low level of, HER2 expression and gene amplification.
Biomarker-stratified “Enrichment” Strategy

• Trial population consists of both biomarker+ and biomarker- patients

• Appropriate design option if
  • treatment is more likely to be effective in the biomarker+ patients, but
  • a clinical meaningful effect could not be ruled out in biomarker- patients

• All-Comer design by enrolling both biomarker+ and biomarker- patients in stratified fashion.
  • Same sample size requirement for the biomarker+ subgroup to that of an enrichment design
Biomarker-stratified Enrichment Strategy: Analysis

- Two main analysis approaches for such enrichment design
  - Parallel testing approach
  - Sequential testing approach
- Parallel testing approach: alpha splitting
  - Test biomarker+ and biomarker- subgroups simultaneously.
  - Success of one subgroup does not depend on the success of the other subgroup
  - Larger sample size requirement compares to full alpha (α) spending approach
  - Could make statistical significance harder to achieve in the biomarker+ subgroup
Biomarker-stratified Enrichment Strategy: Analysis

• Sequential testing approach 1
  • Biomarker sequential testing approach
  • Biomarker+ subgroup is tested first \( \alpha_1 \) (reduced \( \alpha \))
  • If significant, biomarker- subgroup is tested at \( \alpha \)
  • If not significant, the overall population is tested at \( \alpha - \alpha_1 \)

• Sequential testing approach 2
  • Overall followed by biomarker+ testing approach
  • Overall population is tested first at \( \alpha_1 \) (a reduced \( \alpha \))
  • If significant, no further testing is needed
  • If not significant, biomarker+ subgroup is tested at \( \alpha - \alpha_1 \).
Biomarker-stratified Enrichment Strategy: Analysis

- Sequential testing approaches 1 & 2
  - Larger sample size requirement for both overall and biomarker+ subgroup
  - A reduced $\alpha$ could make statistical significance harder to achieve
    - in both the biomarker+ subgroup and overall population

- Sequential testing approach 3
  - Biomarker+ subgroup followed by overall population
  - Allows the use of full alpha for both tests

- Biomarker-stratified enrichment requires a larger study
  - Biomarker+ patients is small portion of overall population
Hypothetical Example

- Two arm study: 10% biomarker-positive subpopulation

- 468 biomarker+ patients are needed
  - 90% power to detect $\Delta$ of 1.5 with $\sigma$ of 5.0 at $\alpha = 0.05$

- Biomarker+ only enrichment design: $n = 468$
  - Loss of opportunity for studying biomarker- subgroup and thus overall population

- Biomarker-stratified enrichment design: $n = 4680$
  - Large $n$ due to the required number of Biomarker+ subjects
  - Hugely overpowered if treatment effect is homogeneous across the biomarker-defined subgroups
  - Costly study if no treatment effect in biomarker- patients
Proposed Enrichment Design

• Enrichment design with patient population augmentation
  • Enrich biomarker+ subgroup
  • n = 852: 468 biomarker+ and 384 biomarker- patients

• Sequential testing strategy:
  • Test of biomarker+ subgroup first at full $\alpha$
  • If significant, test of overall population at full $\alpha$

• How do we assess treatment effect for overall population in such enrichment design setting?
Proposed Enrichment Design: Overall Population

- Naïve pooling of biomarker+ and biomarker- patients does not address relevant question
  - Result will be driven by biomarker+ subgroup

- Assume $\delta_+$ and $\delta_-$ are true treatment effects in biomarker+ and biomarker- subgroups respectively

- Treatment effect in overall population can be expressed as weighted average

\[ \delta = p \delta_+ + (1-p) \delta_- \]

- $p$ is prevalence rate of biomarker+ subpopulation
Proposed Enrichment Design: Overall Population

• Estimator of $\delta$

$$\hat{\delta} = \hat{p} \, \hat{\delta}_+ + (1 - \hat{p}) \, \hat{\delta}_-$$

• $\hat{\delta}_+$ and $\hat{\delta}_-$ are estimators of treatment effects in biomarker+ and biomarker- subgroups respectively

• $\hat{p}$ is the proportion of biomarker+ patients obtained at screening phase
  • Info from screening data, so independent of treatment effects
  • Unbiased estimator of $p$
Statistical Properties of Weighted Estimator: $\hat{\delta}$

- **Unbiased estimate of $\delta$**

  \[
  E(\hat{\delta}) = E\left(E(\hat{\delta} | \hat{p})\right) = p \cdot \delta_+ + (1-p) \cdot \delta_- = \delta
  \]

- **Variance of $\hat{\delta}$**

  \[
  \text{Var}(\hat{\delta}) = E[\text{Var}(\hat{\delta} | \hat{p}) + \text{Var}(E(\hat{\delta} | \hat{p}))]
  = E\left[\hat{p}^2 \frac{2\sigma_+^2}{n_+} + (1-\hat{p})^2 \frac{2\sigma_-^2}{n_-}\right] + \text{Var}[p\delta_+ + (1-\hat{p})\delta_-]
  = (p^2 + \frac{p(1-p)}{m}) \frac{2\sigma_+^2}{n_+} + ((1-p)^2 + \frac{p(1-p)}{m}) \frac{2\sigma_-^2}{n_-}
  + \frac{p(1-p)}{m} (\delta_+ - \delta_-)^2
  \]

  - $n_+ / n_-$ and $\sigma_+ / \sigma_-$: sample size and SD in biomarker+/− patients
  - $m$: screening sample size needed to estimate $\hat{p}$
Hypothetical Example Revisit

- Two arm study: 10% biomarker+ subpopulation
- Study size of 852 biomarker+/- patients
- 468 biomarker+ patients to achieve 90% power to detect $\Delta$ of 1.5 with $\sigma$ of 5.0 at $\alpha = 0.05$
- Additional 384 biomarker- patients to achieve 90% power for overall population
  - Assuming $\Delta$ of 1.5 with $\sigma$ of 5.0 for biomarker- patients
Hypothesis Testing: Sequential Testing

• Step I: testing for $\delta_+$ in biomarker+ subgroup
  • Test statistic using $\hat{\delta}_+$ and $\text{var} (\hat{\delta}_+)$
  • Establishment of efficacy in biomarker+ subgroup

• Step II: testing for $\delta$ in overall population only if step I is successful
  • Test statistic using $\hat{\delta}$ and $\text{var} (\hat{\delta})$
  • Establishment of efficacy in overall population
Comparison of Different Design Approaches

• All-Comer targeting overall population: n = 468
  • Most efficient design when homogeneous treatment effect across the biomarker-defined subgroups
  • Could fail if biomarker- subgroup has smaller or no treatment effect

• Biomarker-stratified enrichment: n = 4680
  • High POS for biomarker+ subgroup
  • Overpowered for overall population

• Enrichment with patient population augmentation: n = 852
  • High POS for biomarker+ subgroup
  • Adequately powered for overall population
Proposed Enrichment Design: Sample Size

<table>
<thead>
<tr>
<th>p</th>
<th>Enrichment (n_+)* (marker+ only)</th>
<th>Power of Enrichment design for overall (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>90%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>80%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>70%</td>
</tr>
<tr>
<td>10%</td>
<td>468</td>
<td>852</td>
</tr>
<tr>
<td></td>
<td></td>
<td>754</td>
</tr>
<tr>
<td></td>
<td></td>
<td>692</td>
</tr>
<tr>
<td>30%</td>
<td>468</td>
<td>720</td>
</tr>
<tr>
<td></td>
<td></td>
<td>652</td>
</tr>
<tr>
<td></td>
<td></td>
<td>610</td>
</tr>
<tr>
<td>50%</td>
<td>468</td>
<td>624</td>
</tr>
<tr>
<td></td>
<td></td>
<td>576</td>
</tr>
<tr>
<td></td>
<td></td>
<td>550</td>
</tr>
</tbody>
</table>

* Sample size required to achieve 90% power
Design of Enrichment Trial with Patient Population Augmentation

• Adequate power for biomarker+ sub
  • 80% or 90% power targeting on δ+
  • Guarantee high POS for biomarker+ subgroup

• Appropriate power for overall population
  • Maintain reasonable POS for overall population
  • Sample size calculation for overall population based on weighted estimator to achieve reasonable power, say at least at least 70% power
Simulation Setting

- Treatment effect in biomarker+ patients is 1.5
- Treatment effect in biomarker- patients is 1.0, 1.2, 1.4
- Prevalence rate of marker+ patients is 10%, 30%, 50%
## Simulation Results: Study Size

<table>
<thead>
<tr>
<th>p</th>
<th>δ</th>
<th>All-comer (n)</th>
<th>Biomarker-stratified Enrichment (n)</th>
<th>Proposed Enrichment design</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Biomarker+ (n&lt;sub&gt;+&lt;/sub&gt;)</td>
</tr>
<tr>
<td>10%</td>
<td>1.0</td>
<td>954</td>
<td>4680</td>
<td>468</td>
</tr>
<tr>
<td></td>
<td>1.2</td>
<td>696</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.4</td>
<td>530</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30%</td>
<td>1.0</td>
<td>796</td>
<td>4680</td>
<td>468</td>
</tr>
<tr>
<td></td>
<td>1.2</td>
<td>632</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.4</td>
<td>514</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50%</td>
<td>1.0</td>
<td>674</td>
<td>4680</td>
<td>468</td>
</tr>
<tr>
<td></td>
<td>1.2</td>
<td>578</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.4</td>
<td>500</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Sample size required to achieve 90% power
Efficiency of Weighted Estimator

- Fixed total sample size of $n = n_+ + n_-$ per arm, the optimal ratio to achieve the smallest variance

$$\frac{n_+}{n_-} = \sqrt{\frac{(p^2 + \frac{p(1-p)}{m})\sigma_+^2}{((1-p)^2 + \frac{p(1-p)}{m})\sigma_-^2}}$$

- Most efficient when the trial population constitution matches with the disease prevalence

$$\frac{n_+}{n_-} = \frac{p}{(1-p)}$$

as $p(1-p)/m \to 0$ and $\sigma_+ = \sigma_-$
Conclusion

• Weighted test statistic provides a statistical approach to assess overall treatment effect
  • Screening data from the trial is utilized to define the weight

• Proposed enrichment design allows assessment of treatment effects in both biomarker+ subpopulation and overall population while protecting the type-1 error rate

• Proposed enrichment design requires modestly larger n than all-comer design, but trial POS is increased by securing the success of biomarker+ subpopulation
  • The larger the biomarker+ prevalence rate, the fewer the required biomarker- subjects
Conclusion

• Variability of the weighted test statistic is dependent of the sample size of the biomarker-subgroup
  • The less the biomarker-subgroup, the more the variability of weighted test statistic, hence the less power for the overall population based on the weighted test statistic

• If biomarker- is one of the subpopulation of interest, one should plan to have enough such patients for safety database


