A GENERALIZED FRAMEWORK OF OPTIMAL TWO-STAGE DESIGNS FOR EXPLORATORY BASKET TRIALS

Heng Zhou
Early Oncology Statistics, Merck

ASA NJ & PT Chapters Annual Symposium
June 23rd, 2023
Background
As of Dec 2021, there are 5,683 clinical trials assessing anti-PD1/PDL1 mAbs – as monotherapy or in combination with other treatments; 278% increase in the past 5 years (Upadhaya et al. 2022, *Nature Reviews Drug Discovery*)
Clinical Trials Facilitation and Coordination Group

CTFG

Recommendation Paper on the Initiation and Conduct of Complex Clinical Trials

12 February 2019
Basket Trials

- FDA has broad definitions on basket trials.

A master protocol designed to test a single investigational drug or drug combination in different populations defined by disease stage, histology, number of prior therapies, genetic or other biomarkers, or demographic characteristics is commonly referred to as a basket trial (shown in Figure 1).

Figure 1: Schematic Representation of a Master Protocol With Basket Trial Design

* T = investigational drug; D = protocol defined subpopulation in multiple disease subtypes.
Basket Trials

• Basket trials are most widely designed to test the treatment effect of a drug on different indications.

• Purpose: to identify “active” indications to the test drug.

• Endpoint: overall response rate (ORR) in the exploratory phase.

Does the drug work in any of the indications?
Hypothetical Outcome of a Basket Trial

- Five tumor cohorts (n=25 each) in patients refractory to PD-1 treatment (ORR under null: 10%)
- Number of responses range from 2 (8%) to 6 (24%)
• Each tumor cohort is evaluated separately, with or without multiplicity adjustment
Ad-hoc Assessment

- **Clinical director 1**: Look at the 3 top ones! The drug is working!!
- **Clinical director 2**: This is cherry-picking!

ORR under null: 10%
Information Borrowing

• Pooling all the indications to conduct analysis.
  – Too extreme information borrowing.
  – Treatment effect from active indications will be diluted by inactive ones.

• Bayesian information borrowing
  – Thall et al. 2003, Berry et al. 2013, Simon et al., 2016, Cunanan et al., 2017
  – Too complicated in practice
  – Not robust under small sample size with inflated type I error

Is there a simple and robust approach to “pick the cherries”?
• Prune “inactive” indications first and then conduct pooling analysis on the rest indications.

• Penalty adjustment will be paid on the pooling analysis for the possible erroneous pooling.
Two-stage Designs with Pruning & Pooling
Design Overview

• Test if the drug is effective on at least one indication
  – Global null hypothesis: the test drug is ineffective on any indications

• Flexibility: investigators are allowed to specify null/alternative response rates for different indications

• A natural extension of Simon’s optimal two-stage design from one-arm to multi-arms
  – Type I and II error rates explicitly controlled
  – Minimize the expected sample size under null hypothesis or minimize the maximum sample size

• Using pruning & pooling approach
  – Prune inactive indications in stage 1 and conduct pooling analysis in stage 2 on the rest indications
Input Parameters

• $K$: number of tumor indications in the basket trial

• $\mathbf{p}_0 = (p_{10}, \ldots, p_{K0}), \mathbf{p}_1 = (p_{11}, \ldots, p_{K1})$: probability vectors for null and alternative hypothesis
  
  – $H_{k0}: p_k = p_{k0}, H_{k1}: p_k = p_{k1}, k = 1, \ldots, K$; ($p_k$: true response rate of the $k^{th}$ indication)

• $\alpha$: global type I error level

• $\beta$: expected overall type II error level
Design Parameters

- \( n_1 = (n_{11}, \ldots, n_{K1}), n_2 = (n_{12}, \ldots, n_{K2}) \): number of patients enrolled in each indication in stage 1 and stage 2

- \( N = (N_1 = n_{11} + n_{12}, \ldots, N_K = n_{K1} + n_{K2}) \): maximum sample size

- \( r = (r_1, \ldots, r_K) \): pruning bar in stage 1
  - The \( k^{th} \) indication will be pruned if the number of responses is less than \( r_k \)

- \( \alpha^* \): critical value of pooling analysis
Trial Example 1

Stage 1

T1 ($N_1=38$) 5% vs 20%
- $T_1 (n_{11} = 7)$ $r_1 = 1$
  - # response: 1

T2 ($N_2=47$) 10% vs 25%
- $T_2 (n_{21} = 9)$ $r_2 = 2$
  - # response: 3

T3 ($N_3 =60$) 20% vs 35%
- $T_3 (n_{31} = 11)$ $r_3 = 4$
  - # response: 1

Stage 2

T1 ($n_{12} = 31$) # response: 3

T2 ($n_{22} = 38$) # response: 4

Pruned

Decision Criterion: P value<0.041 ($\alpha^*$)

Pooled Tumor Negative

Pooled Tumor Positive

Yes

No

16
Probability of Rejecting Global Null

- $X_{k1}, X_{k2}$: number of responses in stage 1 and stage 2 for indication $k$.
- $m = (m_1, \ldots, m_K)$: pooling indicator, where $m_k = 1_{X_{k1} \geq r_k}; M = \sum m_k$.
- $g = (g_1, \ldots, g_K)$: active indicator, where $g_k = 1$ means indication $k$ is active; 0 otherwise.

- Probability of rejecting global null

$$F(r, n_1, n_2, \alpha^*, p_0, p_1 | g, m) = \prod_{k=1}^{K} \left\{ [B_{1k}^{1-m_k}(1 - B_{0k}^{m_k})]g_k [B_{0k}^{1-m_k}(1 - B_{0k}^{m_k})]^{1-g_k} \right\} \times \sum_{x_{11}=r_1}^{n_{11}} \cdots \sum_{x_{M1}=r_M}^{n_{M1}} \{ \text{Pr}(X_{k1} = x_{k1}, k = 1, \ldots, M) \times \text{Pr}(\sum_{k=1}^{M} (X_{k1} + X_{k2}) > R_M(\alpha^*)) \}$$

$B_{0k} \sim \text{Binom}(r_k - 1; n_k, p_{k0}), B_{1k} \sim \text{Binom}(r_k - 1; n_k, p_{k1})$. 

Stage 1: probability of pruning & pooling

Stage 2: probability of pooling analysis being significant
Type I Error Rate

• Under global null (no treatment effects on any indications): $g = (0, \cdots, 0)$

• Type I error rate:

$$\alpha = \sum_{\{m : \sum m_k \geq 1\}} F(r, n_1, n_2, \alpha^*, p_0, p_1 | g = (0, \cdots, 0), m)$$

• Solve $\alpha^*$ given global type I error level $\alpha$. 

Type II Error Rate

• Given $G = \sum g_k$ active indications and $K - G$ inactive indications

• When $G \geq 1$, the power of design is:

$$1 - \beta(G) = \frac{1}{\text{Card} \left( \{ g : \sum g_k = G \} \right)} \sum_{\{ g : \sum g_k = G \}} \sum_{\{ m : \sum m_k \geq 1 \}} F(r, n_1, n_2, \alpha^*, p_0, p_1 | g, m)$$

• Under the non-informative uniform assumption on the number of truly active indications, the overall type II error rate is:

$$\beta = \frac{1}{K} \sum_{G=1}^{K} \beta(G)$$

Probability of rejecting global null
Optimize Design Parameters

- \( r, n_1, N, \alpha^* \): design parameters need to be optimized

1. \( r, n_1, N \)
2. Global type I error \( \leq \alpha \)
   - Solve \( \alpha^* \)
   - Expected overall power \( \geq 1 - \beta \)
3. Minimize expected sample size under the null or the total sample size

Optimization criteria of Simon’s two-stage design
Sample Size Calculation

• The expected sample size under the null hypothesis

\[ EN(r, n_1, n_2, \alpha^*, p_0) = \sum_{k=1}^{K} n_{k2} \Pr(X_{k1} \geq r_k) + \sum_{k=1}^{K} n_{k1} \]

\[ = \sum_{k=1}^{K} n_{k2} (1 - B(r_{k1} - 1; n_{k1}, p_{k0})) + \sum_{k=1}^{K} n_{k1} \]

• Closed-form sample size!
## Examples of Optimized Design Parameters

For $K = 4$

<table>
<thead>
<tr>
<th>$p_0$ (%)</th>
<th>$N$</th>
<th>$n_1$</th>
<th>$r$</th>
<th>$\alpha^*$ (%)</th>
<th>EN</th>
</tr>
</thead>
<tbody>
<tr>
<td>(5, 5, 20, 20)</td>
<td>(30, 30, 52, 52)</td>
<td>(7, 7, 11, 11)</td>
<td>(1,1,4,4)</td>
<td>3.2</td>
<td>63</td>
</tr>
</tbody>
</table>

For $K = 6$

<table>
<thead>
<tr>
<th>$p_0$ (%)</th>
<th>$N$</th>
<th>$n_1$</th>
<th>$r$</th>
<th>$\alpha^*$ (%)</th>
<th>EN</th>
</tr>
</thead>
<tbody>
<tr>
<td>(5, 5,10, 10, 20, 20)</td>
<td>(29, 29, 37, 37, 49, 49)</td>
<td>(5, 5, 7, 7, 10, 10)</td>
<td>(1,1,2,2,4,4)</td>
<td>2.6</td>
<td>73</td>
</tr>
<tr>
<td>(5, 5, 5, 20, 20, 20)</td>
<td>(27, 27, 27, 49, 49, 49)</td>
<td>(5, 5, 5, 10, 10, 10)</td>
<td>(1,1,1,4,4,4)</td>
<td>2.5</td>
<td>74</td>
</tr>
</tbody>
</table>

For $K = 8$

<table>
<thead>
<tr>
<th>$p_0$ (%)</th>
<th>$N$</th>
<th>$n_1$</th>
<th>$r$</th>
<th>$\alpha^*$ (%)</th>
<th>EN</th>
</tr>
</thead>
<tbody>
<tr>
<td>(5, 5,5,10, 10, 20, 20, 20)</td>
<td>(24, 24, 24, 29, 29, 37, 37, 37)</td>
<td>(4, 4, 4, 8, 8, 8, 8, 8)</td>
<td>(1, 1, 1, 2, 2, 3, 3, 3)</td>
<td>1.6</td>
<td>89</td>
</tr>
</tbody>
</table>

- Target 15% improvement in alternative response rates
Trial Example 2: with fixed budget

Stage 1

- **T1** \((N_1 = 30)\)
  - 5% vs 20%
  - \(n_{11} = 18\)
  - \(r_1 = 2\)
  - # response: 2

- **T2** \((N_2 = 38)\)
  - 10% vs 25%
  - \(n_{21} = 24\)
  - \(r_2 = 4\)
  - # response: 5

- **T3** \((N_3 = 52)\)
  - 20% vs 35%
  - \(n_{31} = 32\)
  - \(r_3 = 9\)
  - # response: 1

Stage 2

- **T1** \((n_{12} = 12 + 10)\)
  - # response: 3

- **T2** \((n_{22} = 14 + 10)\)
  - # response: 4

Reallocate Stage 2 sample size \((n = 20)\) of T3

**Decision Criterion:**
- \(P\) value < 0.031 \((\alpha^*)\)

**Pooled Tumor Negative**
- Yes
- No

**Pooled Tumor Positive**
- 10
- 10
- 20

Pruned
Numerical Study
Hypothetical Trial Settings

- Consider $K = 6$ indications;
- Null response rates: $p_0 = (0.05, 0.05, 0.05, 0.2, 0.2, 0.2)$;
- Alternative response rates: $p_1 = (0.2, 0.2, 0.2, 0.35, 0.35, 0.35)$;
- Controlled type I error level $\alpha = 0.05$; type II error level $\beta = 0.20$;
- Optimized design parameters:
  - $N = (27, 27, 27, 49, 49, 49)$
  - $n_1 = (5, 5, 5, 10, 10, 10)$
  - $r = (1, 1, 1, 4, 4, 4)$
- 10,000 simulated trials.
Performance Metrics

• Probability of claiming the drug works, which is defined as the percentage of the simulated trials in which the drug was claimed as effective in at least one indication.

• Probability of identifying at least two true positives, which is defined as the percentage of the simulated trials in which the drug was claimed as effective in at least 2 truly active indications.

• The expected number of true positives, which is defined as the average number of active indications correctly identified as active in the simulated trials.

• The expected number of false positives, which is defined as the average number of inactive indications incorrectly identified as active in the simulated trials.
Power of claiming positive

Independent Pool
Simon’s Bayesian Prune & Pool
True/False positives

- Independent Pool
- Simon’s Bayesian Prune & Pool

Expected Number vs. True Positive and False Positive
Design with Aggregated Futility Analysis
Motivation

• In a basket trial, tumor cohorts usually have different enrollment speed.
• Current practice is to perform interim futility analysis separately for each cohort once a pre-specific number of patients are enrolled.

How to make the futility decision earlier with available data from all tumor cohorts?
Proposed Solution – Aggregated Futility Analysis

One futility analysis across all cohorts
The total sample size across all cohorts is pre-specified for the futility analysis, while the sample size per cohort is unspecified and flexible.

Conduct one futility analysis by pooling all tumor indications and making one futility decision across all tumor cohorts.

Use pruning and pooling method for the final analysis.
Example of Design with Aggregated Futility Analysis

$H_0: p_k = 5\%; \quad H_1: p_k = 20\%$

### Stage I (Pooling)

- **T1** ($N_1 = 21$)
  - $T1 (n_{11})$
  - $X_{11} = 1$
- **T2** ($N_2 = 21$)
  - $T2 (n_{12})$
  - $X_{12} = 2$
- **T3** ($N_3 = 21$)
  - $T3 (n_{13})$
  - $X_{13} = 1$

**Pooling:**

- **P value < 0.4 ($\alpha_1$)**
  - Yes
  - Claim positive in pooled tumors

**Claim no effect; Stop the trial.**

$n_{11} + n_{12} + n_{13} = 27 \ (S)$

### Stage II (Pruning and Pooling)

- **T1** ($n_{21}$)
  - $T1 (n_{21})$
  - $X_{21} = 1$
  - $r_1 = 2$
  - **Yes**
  - **Pruned**
- **T2** ($n_{22}$)
  - $T2 (n_{22})$
  - $X_{22} = 3$
  - $r_2 = 2$
- **T3** ($n_{23}$)
  - $T3 (n_{23})$
  - $X_{23} = 4$
  - $r_3 = 2$

**Pooling:**

- **P value < 0.019 ($\alpha_2$)**
  - Yes
  - Claim positive in pooled tumors

**Claim no effect**
Compared with Optimal Design

Four tumor cohorts (K=4) under Homogeneous Setting

<table>
<thead>
<tr>
<th>$p_0$</th>
<th>$p_1$</th>
<th>Design with aggregated futility analysis</th>
<th>Optimal design (individual futility analysis)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>SS (Futility Analysis)</td>
<td>Total SS</td>
</tr>
<tr>
<td>1%</td>
<td>15%</td>
<td>23</td>
<td>44</td>
</tr>
<tr>
<td>5%</td>
<td>20%</td>
<td>37</td>
<td>84</td>
</tr>
<tr>
<td>10%</td>
<td>25%</td>
<td>47</td>
<td>128</td>
</tr>
</tbody>
</table>

Both designs control global type-I error at 0.05 and target expected power at 0.8.

Design with aggregated futility analysis tends to have less total SS, though the SS for futility analysis and expected SS under null could be slightly larger.
Summary
Take-home messages

• The optimal two-stage basket trial design is a natural extension of Simon’s optimal two-stage design from one-arm to multi-arms.

• To allow more flexibility, we can consider two-stage design with aggregated futility analysis in the first stage.

• The proposed designs are straightforward to implement and have good and comparable operating characteristics as other information borrowing approaches.
• Confirmatory trials with different types of endpoints, e.g., continuous, time-to-event, can also be optimized similarly.

• More patients may be enrolled to confirm the initial findings as inactive (or less active) indications may be included in the pooled analysis.

• Benefit of finding an active new drug often outweighs the risk of wrong tumor selection
  – Additional investigation follows only if drug is deemed active with confidence

• Additional evidence may be necessary to decide on which exact indications to expand cohort to large-scale confirmatory studies, and a risk-mitigated approach may be considered in case of uncertainty.


• Expansion Cohorts: Use in First-in-Human Clinical Trials to Expedite Development of Oncology Drugs and Biologics Guidance for Industry.

• Master protocols: Efficient Clinical Trial Design strategies to Expedite Development of Oncology Drugs and Biologics. FDA Guidance for Industry.