Increasing Efficiency of Oncology Basket Trials using Bayesian Approach

Rong Liu
Pharmaceuticals Statistics, Bayer HealthCare Pharmaceuticals Inc., Whippany NJ

4th ANNUAL ASA NJ CHAPTER/BAYER STATISTICS WORKSHOP
11/11/2016
Acknowledgment

• Co-authors on the present work:
  o Alex Liu (UT Health Science Center at Houston)
  o Mercedeh Ghadessi (Bayer)
  o Richard Vonk (Bayer)

• We appreciate the encouragement and support from Sylvia Engelen and Daniel Haverstock from Bayer on this work
Agenda

- Introduction
  - Cancer Drug Development
  - Basket Trial Design
  - Bayesian Hierarchical Modeling
- Proposed Novel Design for Basket Trial using Bayesian Hierarchical Mixture Modeling
- Simulation Studies
- Summary
Cancer Drug Development Overview

- For Proof of Concept (POC) study, usually single arm small-scale studies to detect efficacy signal and are evaluated based on clinical and imaging criteria such as response rate (RR).
- Cancer is a disease that has been characterized and investigated separately based on the anatomic location: **More than 200 different types of cancer are determined based on the anatomic location!**

---

1. [1] Bayer HealthCare
Targeted Therapy in Oncology

- In last decade, researchers have realized that majority of cancers have genetic risk factors
  - BCR-ABL translocation, two chromosomes switch places (9 and 22)
  - Results in a “fusion gene” created by dis-positioning on ABL and BCR (BCR-ABL Cancers)
Targeted Therapy in Oncology

- **BCR-ABL Cancers** can be found in multiple cancer types
  - Chronic myeloid leukemia (CML)
  - Gastrointestinal stromal tumor (GIST)
  - Acute lymphoblastic leukemia (ALL)
  - Acute myelogenous leukemia (AML)

- Conducted clinical trial studies separately for CML, GIST, ALL, and AML\(^3-5\)

http://path.svhm.org.au/services/Pages/Cytogenetics.aspx
Basket Trial

What is basket trial⁷?

- Trials based on genomics as opposed to site of origin
- Combing multiple cancer types in a single trial
- Molecular biomarker-selected and molecular subtype is more fundamental than histology
- Identify favorable response with a small number of patients

Answer the questions:

1) Does the treatment work on all studied cancer types?
2) If no, can we identify any cancer types with promising effect?
A targeted therapy focuses on a single genetic aberration and can be effective across multiple cancer types:

- A large number of cancer types can be involved in the aberration
- Low frequency of the aberration
- Rarity of some of the cancers

Basket trials provide an efficient tool to develop targeted cancer therapy!
First Basket Trial: Imatinib (2008)

122 subjects with 40 different malignancies
primary endpoint ORR

Indications sensitive to tyrosin kinases (KIT, PDGFRA, or PDGFRB) inhibitor

Solid tumor
- Synovial sarcoma: 1/16 (6%)
- Aggressive fibromatosis: 2/20 (10%)
- Dermato-fibrosarcoma protuberans: 10/12 (83%)

Hematologic malignance
- Systemic mastocytosis: 1/5 (20%)
- Hyper-eosinophilic syndrome: 6/14 (43%)
- Myelo-proliferative disorder: 4/7 (58%)

- Initially enroll up to 10 patients per cancer type
- Number of patients per indication was not prospectively stipulated
- No power consideration for sample size or inferential methods
- Supplemental indications after pooling from case reports and other trials
Recent Basket Trial: Vemurafenib (2015)

70 subjects with at least 14 different malignancies
primary endpoint ORR

- NSCLC: 8/19 (42%)
- Colorectal Monotherapy: 0/10 (0%)
- Colorectal Monotherapy+ Cetuximab: 1/27 (3%)
- Cholangiocarcinoma: 1/8 (13%)
- ECD/LCH: 6/14 (43%)
- Anaplastic Thyroid Ca: 2/7 (29%)

- An adaptive Simon two stage design was used for all tumor-specific cohorts
- No adjustment was made for multiple hypothesis testing (for false positive findings)
- Allow for additional tumor specific cohorts to be analyzed
- The histologic context is an important determinant of response in BRAF V600–mutated cancers.
- Considered to get FDA approval for these indications
Challenges in Simon Two Stage Design for Basket Trials

- Simon two-stage design
  - Allows for stopping early due to futility
  - Distinguish a clinical meaningful response rate (30%) vs Standard of Care response rate (10%) with 5% type I error and 80% power

- Limitation of Simon two-stage parallel design in basket trials
  - Ignores the commonality among cancer type with same genetics mutation
  - Difficult with rare cancer disease

Can we do better?

CML

<table>
<thead>
<tr>
<th>10 patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;1 response</td>
</tr>
<tr>
<td>≤1 response</td>
</tr>
<tr>
<td>19 more patients</td>
</tr>
</tbody>
</table>

GIST

<table>
<thead>
<tr>
<th>10 patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;1 response</td>
</tr>
<tr>
<td>≤1 response</td>
</tr>
<tr>
<td>19 more patients</td>
</tr>
</tbody>
</table>

ALL

<table>
<thead>
<tr>
<th>10 patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;1 response</td>
</tr>
<tr>
<td>≤1 response</td>
</tr>
<tr>
<td>19 more patients</td>
</tr>
</tbody>
</table>

AML

<table>
<thead>
<tr>
<th>10 patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;1 response</td>
</tr>
<tr>
<td>≤1 response</td>
</tr>
<tr>
<td>19 more patients</td>
</tr>
</tbody>
</table>

- Promising
- Not promising

- Stop

CML

<table>
<thead>
<tr>
<th>10 patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;5 responses</td>
</tr>
<tr>
<td>≤5 responses</td>
</tr>
<tr>
<td>Promising</td>
</tr>
<tr>
<td>Not promising</td>
</tr>
</tbody>
</table>

GIST

<table>
<thead>
<tr>
<th>10 patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;5 responses</td>
</tr>
<tr>
<td>≤5 responses</td>
</tr>
<tr>
<td>Promising</td>
</tr>
<tr>
<td>Not promising</td>
</tr>
</tbody>
</table>

ALL

<table>
<thead>
<tr>
<th>10 patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;5 responses</td>
</tr>
<tr>
<td>≤5 responses</td>
</tr>
<tr>
<td>Promising</td>
</tr>
<tr>
<td>Not promising</td>
</tr>
</tbody>
</table>

AML

<table>
<thead>
<tr>
<th>10 patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;5 responses</td>
</tr>
<tr>
<td>≤5 responses</td>
</tr>
<tr>
<td>Promising</td>
</tr>
<tr>
<td>Not promising</td>
</tr>
</tbody>
</table>
Bayesian Hierarchical Modeling (BHM)

- Hierarchical modeling is a unique methodology that can be used to combine information of different indications.
- Inferences for the parameters not only reflect the information about each indication, via the hierarchical modeling, but also borrow relevant information from other indications.
- Sharing and borrowing information across indications allows exchangeability and improvement of power.

What if there are some indication that is very dissimilar from the rest? (Nugget situation)

- The BHM will give over (or under) parameter estimates and large type one error rate (less power) due to nugget effect.
- An unknown heterogeneity among indications poses a major problem.
Example: Nugget Effect

Observed Response Rates:

\[ p_1 = 0.2, \ p_2 = 0.25, \ p_3 = 0.5, \ p_4 = 0.3 \]

Posterior Estimation BHM, Indications RR

How can we avoid too optimistic/pessimistic borrowing for extreme indications (Nuggets)?
Proposed Design

Novel Design for Basket Trials using Bayesian Hierarchical Mixture Modeling (BHMM)
Proposed Novel Design for Basket Trials using BHMM

**Primary Endpoint:** Evaluate each indication's posterior probability (RR > SOC_k), “Go / No-Go” decision

- Genetic aberration: Cancer A, E, F...
- Assess heterogeneity of treatment effect across baskets using meta-analysis
- Is the basket homogenous?
  - Yes: Considering SOC_k for each ind., calculate Bayesian predictive power
  - No: Run separate indication arm with Simon two-stage design
    - Indication low power?
      - Yes: Resume recruiting
      - No: Bayesian hierarchical mixture model to evaluate posterior prob (RR > SOC_k)

Stage 1
Stage 2
Proposed Design Procedures

• Stage 1
  o Evaluate if response rates are homogeneous across indications
    ▶ Heterogeneous: Simon two stage parallel design
    ▶ Homogeneous: Bayesian predictive power evaluation
  o Apply Bayesian predictive power assessment for early futility rule
    ▶ Non-promising indication, drop the indication
    ▶ Promising indication, move to stage 2

• Stage 2
  o Continue recruiting patients for promising indications
    ▶ Determine “Go/No-go” decision using BHMM
Proposed Design Procedures

**Stage 1**
Initial Enrollment

- Heterogeneity Test
  (P value ≥ 0.20, Homogeneous)

- Bayesian Predictive Power
  (Stop for futility, for example, probability of clinically meaningful (response ≥ SOC) is less than 50%)

**Stage 2**
Resume Enrollment

- Bayesian Hierarchical Mixture Model
  (claim Go / No-Go, using posterior probability)
1. How do we control for nugget situation?

Heterogeneity Test to Mitigate Nugget Situation

- **Meta-analysis random effect model** to test response rate heterogeneity\(^\text{10}\)
  - i. Test extreme low or high response rate indication
  - ii. Specific to binomial data and allows computation on exact binomial test

- **Under logistic-normal random effects model,**
  - Using maximum likelihood procedure, estimated between-study variance \(\tau\),
    \[
    \text{logit}(p_i) \sim \text{normal}\ (\mu, \tau) \quad \Rightarrow \quad p_i = i^{th\ \text{indication\ response\ rate}}
    \]

- **Test for Heterogeneity using Cochran’s Q test\(^{12}\),**
  \[
  H_0: p_i = p \quad \text{VS} \quad H_\alpha: \text{At least one response rate is different}
  
  Q = \sum_{i=1}^{k} \frac{(\hat{p}_i - \hat{p})^2}{\tau} \quad \text{where} \quad \hat{p} = \frac{\sum_{i=1}^{k} (\hat{p}_i)}{k}
  \]
  The test is conducted by comparing Q statistics to a \(\chi_{k-1}^2\) distribution

- **Decision:** If we detect heterogeneity across indications, we recommend to apply Simon two stage parallel design for each indication
Heterogeneity Test to Mitigate Nugget Situation

1. How do we control for nugget situation?

Tuning alpha to achieve reasonable power to detect heterogeneity
Bayesian Hierarchical Model (BHM)

Number of response:

\[ r_i \sim \text{Binomial} \left( n_i, p_i \right), \quad i = 1, \ldots, k \]

\[ \text{logit}(p_i) = \theta_i \]

First stage prior:

\[ \theta_1, \ldots, \theta_k | \mu, \tau \sim N(\mu, \tau) \]

Second stage prior:

\[ \mu \sim N(M, S), \quad \tau \sim \text{InverseGamma}(\alpha, \beta) \]

\( \mu \) represents indication treatment effects;

\( \tau \) represents variation and borrowing strength,

\( \tau = 0 \) corresponds to pooling, large \( \tau \) indicate separate analyses

Can we make the prior more robust?

2. How can we borrow information from similar indications?
3. Can we propose a model that is robust to the prior selection?

**Bayesian Hierarchical Mixture Model (BHMM)**

**Mixture Prior:**
- Heavy tailed mixture distribution is a robust prior\(^{15, 17, 18}\)
- Gives more weight to the data when the data and the prior disagree\(^9\)
- Share more information with observed data when they are similar. Thus achieving high precision for posterior estimation

**Number of response:**
\[
\begin{align*}
  r_i &\sim \text{Binomial} \left( n_i, p_i \right), \quad i = 1, \ldots, k \\
  \text{logit}(p_i) &= \theta_i \\
  \theta_i &= \pi \cdot T_1 + (1 - \pi) \cdot T_2 \\
end{align*}
\]

\(\pi\), risk for inadequacy of prior information is constant

**First stage prior:**
\[
T_1 | \mu_{11}, \tau_{11}^2 \sim N(\mu_{11}, \tau_{11}^2), \quad T_2 | \mu_{22}, \tau_{22}^2 \sim N(\mu_{22}, \tau_{22}^2)
\]

**Second stage prior:**
\[
\begin{align*}
  \mu_{11} &\sim N(\mu_{100}, \sigma_{100}^2), \quad \tau_{11}^2 \sim IG(\alpha, \beta), \\
  \mu_{22} &\sim N(\mu_{200}, \sigma_{200}^2), \quad \tau_{22}^2 \sim IG(\alpha, \beta)
end{align*}
\]
3. Can we propose a model that is robust to the prior selection?

**Comparison between Non-robust (BHM) versus Robust Prior Model (BHMM)**

- $\theta_i \leq SOC_i$ vs $\theta_i > SOC_i$ for any $i$
- GO is defined as posterior prob($\hat{\theta}_i > SOC_i | data$) > 0.9
- True RR: (0.30, 0.30, 0.20, 0.30), SOC: (0.20, 0.25, 0.15, 0.30)
- Simulation was based on 1000 trials

<table>
<thead>
<tr>
<th>Bayesian Hierarchical Model with non-robust prior</th>
<th>Probability of GO for indication 1</th>
<th>Probability of GO or indication 2</th>
<th>Probability of GO for indication 3</th>
<th>Probability of GO for indication 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu \sim N(0.05, 0.49)$</td>
<td>0.56</td>
<td>0.21</td>
<td>0.38</td>
<td>0.06</td>
</tr>
<tr>
<td>$\mu \sim N(0.2, 0.29)$</td>
<td>0.63</td>
<td>0.25</td>
<td>0.39</td>
<td>0.06</td>
</tr>
<tr>
<td>$\mu \sim N(0.8, 0.42)$</td>
<td>0.66</td>
<td>0.31</td>
<td>0.41</td>
<td>0.08</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Bayesian Hierarchical Mixture Model with robust prior</th>
<th>Probability of GO for indication 1</th>
<th>Probability of GO or indication 2</th>
<th>Probability of GO for indication 3</th>
<th>Probability of GO for indication 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_{11} \sim N(0.2, 0.29); \mu_{22} \sim N(0.1, 0.42)$</td>
<td>0.60</td>
<td>0.25</td>
<td>0.39</td>
<td>0.07</td>
</tr>
<tr>
<td>$\mu_{11} \sim N(0.8, 0.42); \mu_{22} \sim N(0.1, 0.42)$</td>
<td>0.61</td>
<td>0.25</td>
<td>0.39</td>
<td>0.07</td>
</tr>
<tr>
<td>$\mu_{11} \sim N(0.05, 0.49); \mu_{22} \sim N(0.1, 0.42)$</td>
<td>0.61</td>
<td>0.25</td>
<td>0.39</td>
<td>0.07</td>
</tr>
</tbody>
</table>
Simulation Studies

- Scenario One: 4 indications in one Basket
  - Target Response Rate: (0.4, 0.4, 0.4, 0.4)
  - SOC Response Rate: (0.2, 0.2, 0.2, 0.2)

- Scenario Two: 5 indications in one Basket
  - Target Response Rate: (0.4, 0.5, 0.4, 0.4, 0.4)
  - SOC Response Rate: (0.15, 0.25, 0.2, 0.2, 0.15)

- Simulation Steps:
  - Interim Analysis
  - Parameter estimation from BHMM
  - Power and sample size evaluation
Scenario 1

<table>
<thead>
<tr>
<th>True Response Rate</th>
<th>p1</th>
<th>p2</th>
<th>p3</th>
<th>p4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Great</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
</tr>
<tr>
<td>Nugget</td>
<td>0.4</td>
<td>0.2</td>
<td>0.2</td>
<td>0.4</td>
</tr>
<tr>
<td>Null</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SOC Rate</th>
<th>soc1</th>
<th>soc2</th>
<th>soc3</th>
<th>soc4</th>
</tr>
</thead>
<tbody>
<tr>
<td>SOC Equal</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
</tr>
</tbody>
</table>

Simulation study setting: (# of simulated trials=1000, # of tumor indications=4)

1. **Great**: Target and underlying response rate for every indication match and all of them demonstrating a promising effect in comparison with their SOC

2. **Nugget**: Indication 1 and 4 similarly show promising effect, the underlying response rate for indication 2 and 3 is almost as good as SOC

3. **Null**: Every indication has an acceptable response rate but not clinically meaningful in comparison with SOC
Matching with Simon’s Two Stage Design Sample Size

- Simon Two stage requires 148 patients for all four indications running in parallel to reach 80% power
- The interim analysis starts with 11 to 15 patients per indication based on Simon’s two Stage Design interim criterion
Scenario 1: Study Diagram – Homogeneous Branch
Great: True=(0.4, 0.4, 0.4, 0.4) vs. SOC=(0.2, 0.2, 0.2, 0.2)

First stage enrollment 11-15pts per indication
Total enrollment \( n_T = 38 \)
Total simulation trials \( T_{sim} = 1000 \)

Homogeneity
(Trials=779 out of 1000)

Bayesian Predictive Power:
\[ P(\text{response} \geq \text{SOC}) \] > 50%

Resume recruitment,
Bayesian Hierarchal Mixture Model:
\[ P_{\text{Posterior}}(\text{response} \geq \text{SOC}) \] > 95%

First stage recruitment
Meta analysis heterogeneous \( \alpha \)-level: 0.2

Drop tumor indication due to futility

221 out of 1000 trials goes to heterogeneous branch

T1 (779)

T1.pass.first (753)

T1.success (716)

T2 (779)

T2.pass.first (750)

T2.success (715)

T3 (779)

T3.pass.first (747)

T3.success (713)

T4 (779)

T4.pass.first (750)

T4.success (715)
Scenario 1: Study Diagram - Homogeneous Branch - Estimations

Great: True=(0.4, 0.4, 0.4, 0.4) vs. SOC=(0.2, 0.2, 0.2, 0.2)

- Parameter estimation, 90% credible interval, bias, and mean squared error using Bayesian hierarchical mixture model for each indications response rate and overall response rate

<table>
<thead>
<tr>
<th>Bayesian Hierarchical Mixture Model</th>
<th>Estimated RR1</th>
<th>Estimated RR2</th>
<th>Estimated RR3</th>
<th>Estimated RR4</th>
<th>Estimated RR_{overall}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter Estimation True: (0.4, 0.4, 0.4, 0.4)</td>
<td>0.405</td>
<td>0.403</td>
<td>0.401</td>
<td>0.404</td>
<td>0.403</td>
</tr>
<tr>
<td>90% Credible Interval</td>
<td>(0.314, 0.502)</td>
<td>(0.306, 0.501)</td>
<td>(0.307, 0.505)</td>
<td>(0.309, 0.510)</td>
<td>(0.338, 0.469)</td>
</tr>
<tr>
<td>Bias</td>
<td>0.005</td>
<td>0.003</td>
<td>0.001</td>
<td>0.004</td>
<td>0.003</td>
</tr>
<tr>
<td>MSE</td>
<td>0.003</td>
<td>0.004</td>
<td>0.003</td>
<td>0.004</td>
<td>0.002</td>
</tr>
</tbody>
</table>
Scenario 1: Simulation Final Results

% GO in Homogenous and Heterogeneous (Matching Sample Size with Simon’s)

Sample Size Saving (Matching Power with Simon’s)

Simon Design: 80% power and 5% α-level for each indication
Scenario 2

True Response Rate | p1  | p2  | p3  | p4  | p5  |
--- | --- | --- | --- | --- | --- |
Great          | 0.4 | 0.5 | 0.4 | 0.4 | 0.4 |
Nugget         | 0.4 | 0.5 | 0.2 | 0.2 | 0.4 |
Null           | 0.15| 0.25| 0.2 | 0.2 | 0.15|
SOC Rate       | soc1| soc2| soc3| soc4| soc5|
SOC Unequal    | 0.15| 0.25| 0.2 | 0.2 | 0.15|

Simulation study setting: (# of simulated trials=1000, # of tumor indications=5)

1. Allow different SOC response rate across 5 indications

2. **Great**: Target and underlying response rate for every indication match and all of them demonstrating a promising effect in comparison with their SOC

3. **Nugget**: Indication 1, 2 and 5 similarly show promising effect, the underlying response rate for indication 3 and 4 is almost as good as SOC

4. **Null**: Every indication has an acceptable response rate but not clinically meaningful in comparison with SOC
Matching with Simon’s Two Stage Design Sample Size

- Simon Two stage requires 138 patients for all five indications running in parallel to reach 80% power
- The interim analysis starts with 5 to 9 patients per indication based on Simon’s two Stage Design interim criterion
Scenario 2: Study Diagram - Homogeneous Branch - Estimations
Great: True=(0.4, 0.5, 0.4, 0.4, 0.4) vs. SOC=(0.15, 0.25, 0.2, 0.2, 0.1)

- Parameter estimation, 90% credible interval, bias, and mean squared error using Bayesian hierarchical mixture model for each indication response rate and overall response rate

<table>
<thead>
<tr>
<th>Bayesian Hierarchical Mixture Model</th>
<th>Estimated RR1</th>
<th>Estimated RR2</th>
<th>Estimated RR3</th>
<th>Estimated RR4</th>
<th>Estimated RR5</th>
<th>Estimated RR_{overall}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter Estimation True: (0.4, 0.5, 0.4, 0.4, 0.4)</td>
<td>0.412</td>
<td>0.468</td>
<td>0.405</td>
<td>0.405</td>
<td>0.415</td>
<td>0.416</td>
</tr>
<tr>
<td>90% Credible Interval</td>
<td>(0.29, 0.55)</td>
<td>(0.34, 0.63)</td>
<td>(0.29, 0.51)</td>
<td>(0.29, 0.52)</td>
<td>(0.28, 0.55)</td>
<td>(0.33, 0.50)</td>
</tr>
<tr>
<td>Bias</td>
<td>0.012</td>
<td>-0.032</td>
<td>0.005</td>
<td>0.005</td>
<td>0.015</td>
<td>0.001</td>
</tr>
<tr>
<td>MSE</td>
<td>0.004</td>
<td>0.006</td>
<td>0.003</td>
<td>0.003</td>
<td>0.005</td>
<td>0.002</td>
</tr>
</tbody>
</table>
Scenario 2: Simulation Results

% GO in Homogenous and Heterogeneous (Matching Sample Size with Simon’s)

Sample Size Saving (Matching Power with Simon’s)

Simon Design: 80% power and 5% \( \alpha \)-level for each indication
Summary
Summary

- Histology-independent, biomarker-selected basket studies can serve as an efficient tool for developing molecularly targeted cancer therapy
- It allows for detection of early efficacy activities across multiple tumor types simultaneously
- Faster identification of efficacious drugs with fewer patients
Summary

• One challenge in interpreting the results of basket studies is drawing inferences from small numbers of patients

• This calls for innovative and efficient design:
  • The proposed design takes practical aspects of basket trial into consideration
  • It is robust to prior selection and allows dynamic borrowing of information
  • It naturally adjusts more borrowing effect when the indication are consistent and less borrowing when the indications are different
  • It saves sample size comparing to tradition two stage design and improve the efficiency of the trial
References

Thank you

Question?