Basket and platform protocols in full development in oncology

Viktoriya Stalbovskaya, Amy Racine, Anthony D’Amelio, Tomas Haas
Basel
October 06, 2017
Outline

• Introduction to master protocols and motivation
• Novartis ROAR study – example of a basket design
• Novartis melanoma platform design – example of a platform design
  – Statistical methodology
  – Simulation results for the melanoma platform
• Regulatory context
Introduction to Master Protocols

• Increasing interest in performing innovative trials allowing for simultaneous evaluation of multiple treatments in one disease or one treatment in multiple diseases within the same overall trial structure.
• Such designs are referred to as master protocols

Table 1. Types of Master Protocols.

<table>
<thead>
<tr>
<th>Type of Trial</th>
<th>Objective</th>
</tr>
</thead>
<tbody>
<tr>
<td>Umbrella</td>
<td>To study multiple targeted therapies in the context of a single disease</td>
</tr>
<tr>
<td>Basket</td>
<td>To study a single targeted therapy in the context of multiple diseases or disease subtypes</td>
</tr>
<tr>
<td>Platform</td>
<td>To study multiple targeted therapies in the context of a single disease in a perpetual manner, with therapies allowed to enter or leave the platform on the basis of a decision algorithm</td>
</tr>
</tbody>
</table>

From Woodcock and LaVange (NEJM 2017)
Motivation

• Recent advances in oncology drug development have improved progression free and overall survival.

• Studies with a “traditional” design with all comers are becoming less feasible.

• Future studies must consider the disease prevalence, the pace of development of new therapies

• Smaller, shorter, more focused studies in a more narrowly defined disease

Source: Trialtrove® | Pharma Intelligence, September 2017
Pros and Cons of Basket/Platform Designs

**PROS**
- One overarching protocol designed to answer multiple questions.
- Shared trial infrastructure
- Cost and time savings
- Adaptive design features adding/dropping arms response adaptive rando

**CONS**
- Longer timelines for initial set up
- Complex trial logistics & operations
Novartis* ROAR
Example of a basket study

Innovative adaptive basket design with hierarchical Bayesian model employed to compensate for the small sample sizes across selected histologies,

Setting: one treatment, multiple histologies/tumor types

Design: multiple single arm cohorts

Primary endpoint: objective response rate (ORR)

Statistical model: Bayesian hierarchical model - borrows info across cohorts; borrowing in limited sense from histologies that show similar ORRs (shrinkage estimation)

Interim analyses: performed at ~ 12week intervals; enrollment for each cohort may be stopped early for futility or efficacy: decisions based on whether posterior probability that the ORR exceeds its corresponding historical control is sufficiently low or high

Patients with rare cancers and BRAF V600E mutation

All treated with dabrafenib+trametinib

* Designed together with Berry Consultants
# ROAR study – results

## Overall response rate

<table>
<thead>
<tr>
<th>Condition</th>
<th>Observed ORR</th>
<th>95% Exact Binomial CI</th>
<th>Estimated ORR</th>
<th>95% Credible Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anaplastic thyroid cancer (N=15)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Biliary tract cancer (N=18)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WHO Gr 1/2 Glioma (N=6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WHO Gr 3/4 Glioma (N=19)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adenocarcinoma of small intestine (N=2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hairy cell leukemia (N=24)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multiple myeloma (N=6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Results presented at ASCO’17.

**Observed ORR and 95% exact binomial confidence interval**  
**Estimated ORR and 95% credible interval based on Bayesian Hierarchical model**
ROAR study – regulatory feedback

• Bayesian hierarchical model and in particular the concept of borrowing information across cohorts was considered acceptable by the FDA statisticians.

• Eventually, FDA was willing to consider alternate trial designs and methodologies, including the Bayesian hierarchical model, to assess efficacy in tumors that are sufficiently rare as to prohibit traditional methods.

• Swedish HA (MPA) expressed skepticism towards the Bayesian hierarchical design and view this trial as 9 single arm studies.

• Dutch HA (MEB) agreed with selecting specific histologies based on the Bayesian design; however, each histology will be judged individually based on the disease context and efficacy data.
Melanoma Platform Trial

Evaluation of multiple combinations in 2/3 line
Pick the Winner Phase II Design

Phase II Platform
Some combinations available, some to be potentially added later, design open for dynamically testing new cohorts

Evaluate

1st Group Cohorts
Combo 1
Combo 2
Combo 3

Potential Subsequent Cohorts
Combo 4
Combo 5
Combo 6
Combo 7

Statistical decision rule based on observed ORRs

Pick the winner

Possible Accelerated approval

Expand winning cohort

Drop cohort (futility rule) → stop enrollment

Continue cohort enrollment

Adapt randomization probabilities

Randomized Phase III trial

Response Adaptive Randomization

Path to Full Approval

Biostatistics and Pharmacometrics

Melanoma Platform Statistical Considerations | September 19, 2017 | OGDU Unit | Confidential
Statistical Summary

- **Bayesian methodology** used in the probabilistic assessment of efficacy (ORR) and in adaptive decisions making
- Extensive **simulations** performed to fine-tune decisions rules
  - Assessments of ORR against pre-defined thresholds performed in a ‘group sequential’ manner
    - Used both batches and enrollment rates and performed interim assessments driven by either batch enrollment or calendar time
    - Further incorporated enrollment of new arm at “off times”, i.e. arms added in between efficacy assessments
    - Also incorporated minimum # of patients enrolled per arm needed for interim assessments as well as capped the maximum number of patients enrolled per arm
    - Possibility to incorporate statistical tools to compare several ‘winning’ arms
  - Response adaptive randomization implemented
- Investigated platform designs demonstrated **reasonable behavior and operating characteristics**
Statistical Bayesian model

- Patients are assessed in batches of size $n$ per arm
  - Size $n$ determined by enrollment rate,
  - number of months per efficacy assessment
  - Inclusion of any new arms during efficacy assessment

- Uninformative prior for true ORR rates $p_1, p_2, p_3, \ldots, p_k$ in treatment arms 1, 2, $\ldots, k$

- At completion of batch 1 (total $N=n^*T$)
  1. $p_i \sim \text{beta}(y_{1i}+1, n_{1i}-y_{1i}+1)$
  2. Decisions
     - Winner $\text{Prob}(p_i \geq r_W | \text{data}) \geq P_W$
     - Futile $\text{Prob}(p_i \leq r_F | \text{data}) \geq P_F$
     - Continue otherwise

$n = \text{batch size, } T = \text{number of arms available.}$

$r_W$ and $r_F$ are ORR thresholds for success (winner) and futility, resp.
$P_W$ and $P_F$ are probability thresholds for success (winner) and futility, resp.
Incorporation of Response Adaptive Randomization (RAR):

1. For arms which are not clear winners or losers after a batch, calculate

- \( P_{\text{best}_j} = Pr(p_j > p_i, \text{for all } j^\wedge = i, \text{continue}|\text{response}_i) \)

- RAR prob: \( P_j = \frac{(p_{\text{best}_j} \cdot \text{Var}(p_j))}{(n_j+1)} \) (B. Serville and Berry 2016 Clinical trial)

- Incorporate a design which allows for randomization in selecting better performing arms without completely sacrificing balance among the arms
  - Double-adaptive biased coin design (Hu and Zhang 2004, Annals of Statistics)
  - Can control exact randomization with gamma (Use 2 for our simulations)

\[
P_j = \Pr(\text{Trt } j \text{ is assigned }) = \frac{p_j \left( \frac{p_j}{\left( \frac{N_{ji}}{N(i)} \right)} \right)^\gamma}{\sum_{j=1}^{k} p_j \left( \frac{p_j}{\left( \frac{N_{ji}}{N(i)} \right)} \right)^\gamma}, \quad j = 1, \ldots, K, \quad N(i) = \text{total size}
\]

2. Continue to next batch (of size \( n \)) with continuing arms (and new arms added) and with RAR applied (resulting in unequal ‘n’s across arms after completing batch 2)
### Design Parameters & Decision Rules Investigated

Design x (assessment batch n/arm; start with K1 arms, later add K2 arms):

<table>
<thead>
<tr>
<th>ORR threshold</th>
<th>&lt;x1%</th>
<th>x1-x2%</th>
<th>≥x2%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decision rule</td>
<td>Prob(ORR&lt;x1%)&gt;p1%</td>
<td>Prob(ORR≥x2%)&gt;p2%</td>
<td></td>
</tr>
<tr>
<td>Action</td>
<td>Drop for futility/ stop enrollment</td>
<td>Continue enrollment</td>
<td>Declare ‘winner’</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Design #</th>
<th># of arms start/add</th>
<th>n/batch</th>
<th>x1</th>
<th>p1</th>
<th>x2</th>
<th>p2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5/2</td>
<td>25</td>
<td>10</td>
<td>80</td>
<td>30</td>
<td>90</td>
</tr>
<tr>
<td>2</td>
<td>5/2</td>
<td>25</td>
<td>15</td>
<td>60</td>
<td>20</td>
<td>60</td>
</tr>
<tr>
<td>3</td>
<td>5/2</td>
<td>20</td>
<td>15</td>
<td>60</td>
<td>20</td>
<td>60</td>
</tr>
<tr>
<td>4</td>
<td>3/1/1</td>
<td>20 with 4 batch total</td>
<td>15</td>
<td>70</td>
<td>20</td>
<td>70</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>8 pts/mth. with 4 batch total</td>
<td>15</td>
<td>70</td>
<td>20</td>
<td>70</td>
</tr>
<tr>
<td>6</td>
<td>4/1/1 (‘off times’)</td>
<td>8 pts/mth. with 6 batch total</td>
<td>15</td>
<td>70</td>
<td>20</td>
<td>70</td>
</tr>
</tbody>
</table>

Compare against standard design

---

Business Use Only - CONFIDENTIAL
**Illustration of Design Operating Characteristics**

**Design 5/Scenario 1: Decision summary and Power**

<table>
<thead>
<tr>
<th>Simulation scenario</th>
<th>4 arms</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>True ORR</td>
<td>5%</td>
</tr>
</tbody>
</table>

- **‘Futile’ arm 1 (5%)**: 54% prob to drop after 1st batch (N = 12), 90% after 2nd batch; avg N=22
- **‘Futile’ arm 2 (7%)**: 39% prob to drop after 1st batch (N=12), 79% after 2nd batch; avg N=26
- **‘Interesting’ arm 3 (15%)**: 60% prob after 1st batch to continue (N = 12), 13% to drop, 26% winner; avg N=30
- **‘Winner’ arm 4 (30%)**: 73% prob after 1st batch to declare winner (power), 25% to continue; after 2nd batch: 89% vs 9%; avg N=19

**Decision rules:**

- Prob(ORR ≤ 15%) ≥ 0.70 → drop for Futility
- Prob(ORR ≥ 20%) ≥ 0.70 → declare Winner

Used to implement our objective to drop combos with ORR<10% and pick winners with ORR≥25%

**Relatively good OC**

Design with 4 arms at start with no additional arms added
Max enroll = 45
Efficacy check every 6 months

- **Batch = 1**
- **Batch = 2**
- **Batch = 3**
- **Batch = 4**
Illustration of Design Operating Characteristics

**Design 6/Scenario 1: Decision summary and Power**

<table>
<thead>
<tr>
<th>Simulation scenario</th>
<th>6 arms</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>True ORR</td>
<td>5%</td>
</tr>
</tbody>
</table>

- **Futility**
  - Arm 1 (5%): 57% prob to drop after 1st batch, 86% after 2nd batch; avg N=18
  - Arm 2 (7%): 43% prob to drop after 1st batch, 73% after 2nd batch; avg N=22

- **Interesting**
  - Arm 3 (15%): 64% prob after 1st batch to continue, 16% to drop, 20% winner; avg N=29

- **Winner**
  - Arm 5 (24%): 53% prob after 2nd batch to declare winner (power), 41% to continue; after 6th batch: 68% vs 25%; avg N=30

**Decision rules:**
- Prob(ORR ≤ 15%) ≥ 0.70 → drop for Futility
- Prob(ORR ≥ 20%) ≥ 0.70 → declare Winner

Used to implement our objective to drop combos with ORR<10% and pick winners with ORR≥25%

Design with 4 arms at start with arms added at 4 and 9 months
Min arm = 10; max arm = 45
Efficacy check every 6 months
Questions to the audience

• What are the main limitations the audience can think would prohibit the sponsors from obtaining regulatory approval based on data from a platform study?

• Any practical experience in benefits of traditional vs platform studies when the operating characteristics are similar (e.g. no borrowing and each group is treated as a separate strata)?

• Any feedback on borrowing information either across indications, combinations, or sharing of common control arm?

• Does the audience foresee regulatory issues with the response-adaptive randomization (RAR) in general and with the RAR in presence of combinations coming and going?

• Any positive precedent of running truly global studies with new drugs being added under master protocol. Any practical recommendations how to keep the amendments light?

• Which benefits and limitations does the audience foresee with the multiple groups to assess safety especially in rare diseases (although it is applicable to basket trials more than platforms)?

• Is there a concern over study integrity if the study design is adapted to reflect evolving treatment landscape (e.g. increase the success and failure bars)?
Acknowledgement

• Emmanuel Zuber
• Alex Sverdlov
• Simon Wandel
References


• Vivek Subbiah et al (2016). ROAR: a phase 2, open-label study in patients (pts) with BRAF V600E–mutated rare cancers to investigate the efficacy and safety of dabrafenib (D) and trametinib (T) combination therapy. // Journal of Clinical Oncology 2016 34:15_suppl, TPS2604-TPS2604

Back-Up
## ROAR study results*

<table>
<thead>
<tr>
<th>Cohort $</th>
<th>Historical Control Response Rate</th>
<th>Number of ITT/Evaluable Subjects</th>
<th>Number of Confirmed Response</th>
<th>Observed Overall Response Rate</th>
<th>Estimated Response Rate and 95% Credible Interval [1]</th>
<th>Prob. that the ORR Exceeds Historical Control Rate [1]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anaplastic thyroid cancer</td>
<td>15%</td>
<td>15</td>
<td>11</td>
<td>73% (44.9, 92.2)</td>
<td>69% (46.9, 86.9)</td>
<td>1.00</td>
</tr>
<tr>
<td>Biliary tract cancer</td>
<td>10%</td>
<td>18</td>
<td>7</td>
<td>39% (17.3, 64.3)</td>
<td>41% (21.3, 62.3)</td>
<td>1.00</td>
</tr>
<tr>
<td>WHO Gr 1/2 Glioma</td>
<td>10%</td>
<td>6</td>
<td>3</td>
<td>50% (11.8, 88.2)</td>
<td>49% (20.2, 76.8)</td>
<td>1.00</td>
</tr>
<tr>
<td>WHO Gr ¾ Glioma</td>
<td>10%</td>
<td>19</td>
<td>4</td>
<td>21% (6.1, 45.6)</td>
<td>27% (8.9, 47.8)</td>
<td>0.96</td>
</tr>
<tr>
<td>Adenocarcinoma of small intestine</td>
<td>10%</td>
<td>2</td>
<td>1</td>
<td>50% (1.3, 98.7)</td>
<td>48% (11.3, 82.0)</td>
<td>0.98</td>
</tr>
<tr>
<td>Hairy cell leukemia</td>
<td>10%</td>
<td>24</td>
<td>18</td>
<td>75% (53.3, 90.2)</td>
<td>69% (50.1, 85.3)</td>
<td>1.00</td>
</tr>
<tr>
<td>Multiple myeloma</td>
<td>15%</td>
<td>6</td>
<td>3</td>
<td>50% (11.8, 88.2)</td>
<td>54% (22.9, 80.9)</td>
<td>0.99</td>
</tr>
</tbody>
</table>

[1] Based on Bayesian Hierarchical model-based analysis
* Interim analysis #8 presented at ASCO 2017; $ There were no subjects enrolled into GIST and NSGCT/NGGCT cohorts.
**Illustration of Sample Sizes**

*Design 5/4 arms at start with no additional arms added*

| Simulation scenario | 4 arms |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |
|---------------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|        |
| True ORR            | 1      | 2      | 3      | 4      |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |
| Scenario 1          | 5%     | 7%     | 15%    | 30%    |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |
| Scenario 2          | 5%     | 7%     | 5%     | 7%     |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |
| Scenario 3          | 5%     | 7%     | 30%    | 35%    |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |
| Scenario 4          | 5%     | 15%    | 18%    | 26%    |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |

### Sample size varies from approximately 21 to 27.

- **‘Futile’ arms**
- Sample sizes higher than clear winners and losers; ‘cap’ incorporated to control max.’N’.

### ‘Interesting’ arms

- Sample sizes vary from 16 to 22. Borderline winners, such as scenario 4 arm 4, will have higher sample sizes.

---

*Sample size in a platform trial is variable and dependent on the true ORRs*

In these scenarios, total average sample size varies from 85-100 patients for these 4 arms.
Illustration of Sample Sizes

Design 6/4 arms at start with arms added at 4 and 9 months

<table>
<thead>
<tr>
<th>Simulation scenario</th>
<th>True ORR</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>Avg. Sample Size</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>Σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scenario 1</td>
<td>5%</td>
<td>7%</td>
<td>15%</td>
<td>30%</td>
<td>24%</td>
<td>19%</td>
<td></td>
<td>18</td>
<td>22</td>
<td>29</td>
<td>17</td>
<td>30</td>
<td>37</td>
<td></td>
<td>153</td>
</tr>
<tr>
<td>Scenario 2</td>
<td>5%</td>
<td>7%</td>
<td>5%</td>
<td>7%</td>
<td>4%</td>
<td>20%</td>
<td></td>
<td>18</td>
<td>22</td>
<td>18</td>
<td>23</td>
<td>23</td>
<td>40</td>
<td></td>
<td>144</td>
</tr>
<tr>
<td>Scenario 3</td>
<td>5%</td>
<td>7%</td>
<td>30%</td>
<td>35%</td>
<td>17%</td>
<td>8%</td>
<td></td>
<td>18</td>
<td>22</td>
<td>18</td>
<td>15</td>
<td>36</td>
<td>35</td>
<td></td>
<td>144</td>
</tr>
<tr>
<td>Scenario 4</td>
<td>5%</td>
<td>15%</td>
<td>18%</td>
<td>26%</td>
<td>33%</td>
<td>7%</td>
<td></td>
<td>18</td>
<td>27</td>
<td>27</td>
<td>21</td>
<td>24</td>
<td>33</td>
<td></td>
<td>150</td>
</tr>
<tr>
<td>Scenario 5</td>
<td>11%</td>
<td>16%</td>
<td>21%</td>
<td>26%</td>
<td>19%</td>
<td>12%</td>
<td></td>
<td>25</td>
<td>26</td>
<td>25</td>
<td>20</td>
<td>33</td>
<td>36</td>
<td></td>
<td>165</td>
</tr>
<tr>
<td>Scenario 6</td>
<td>29%</td>
<td>24%</td>
<td>27%</td>
<td>8%</td>
<td>11%</td>
<td>7%</td>
<td></td>
<td>18</td>
<td>23</td>
<td>20</td>
<td>23</td>
<td>32</td>
<td>32</td>
<td></td>
<td>148</td>
</tr>
</tbody>
</table>

Sample size varies from approximately 18 to 36. Elevated sizes seem to come from the later enrolling arms.

‘Interesting’ arms

Can have sample sizes at least twice as high as clear winners and clear futility arms; and this is especially true for arm 6.

‘Winning’ arms

Sample sizes vary from 15 to 30. Borderline winners, such as scenario 1 arm 5, will have higher sample sizes.

Sample size in a platform trial is highly variable and dependent on the true ORRs

In these scenarios, total average sample size varies from 146-165 patients for these 6 arms