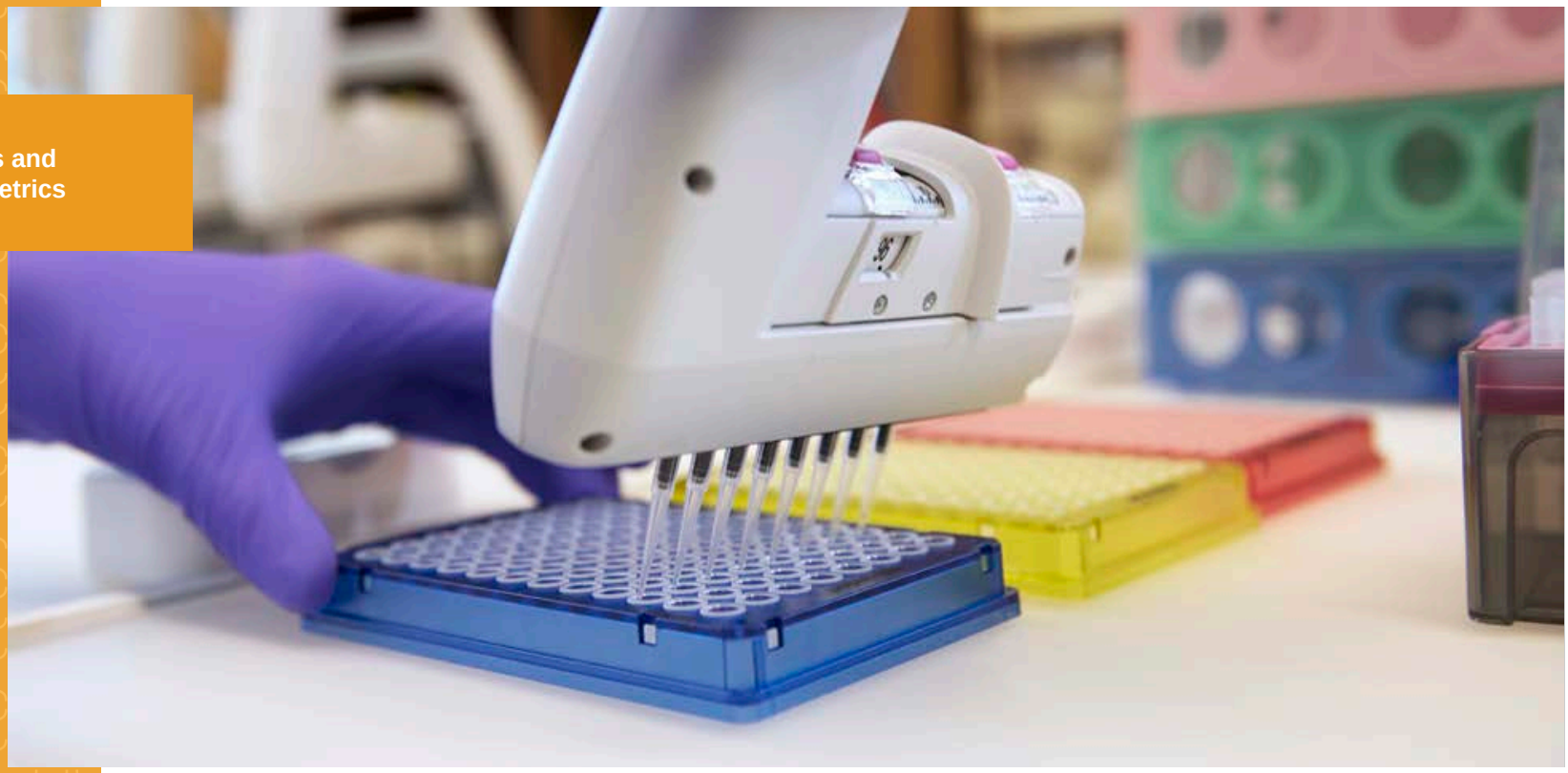
Biostatistics and Pharmacometrics



Basket and platform protocols in full development in oncology

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

NOVARTIS

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.

Type of Trial	Objective
Umbrella	To study multiple targeted therapies in the context of a single disease
Basket	To study a single targeted therapy in the context of multiple diseases or disease subtypes
Platform	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

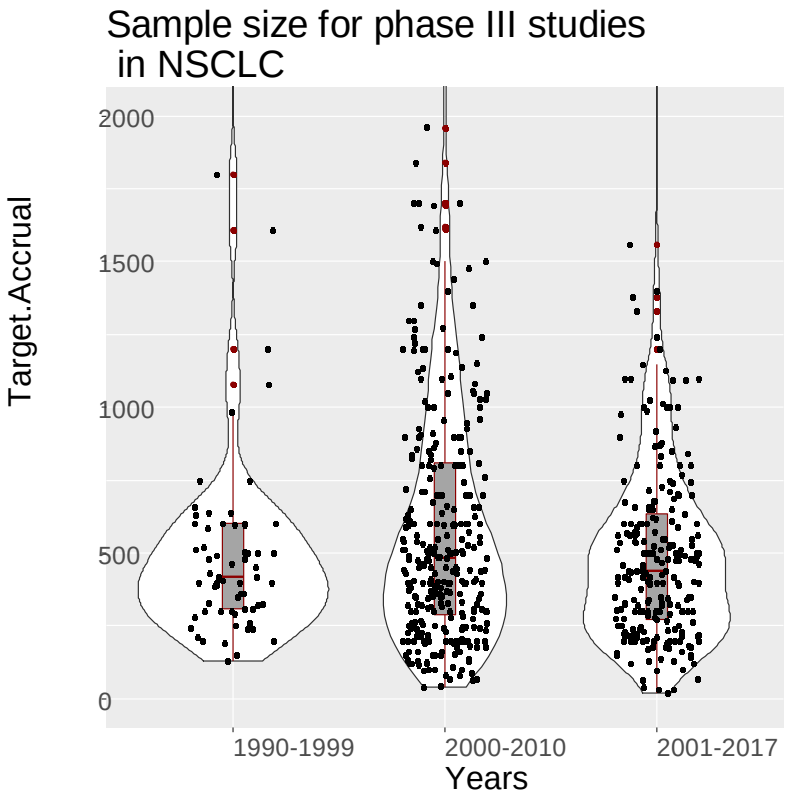
From Woodcock and LaVange (NEJM 2017)

FRACTION-GC
FRACTION-lung

ROAR

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

NOVARTIS

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

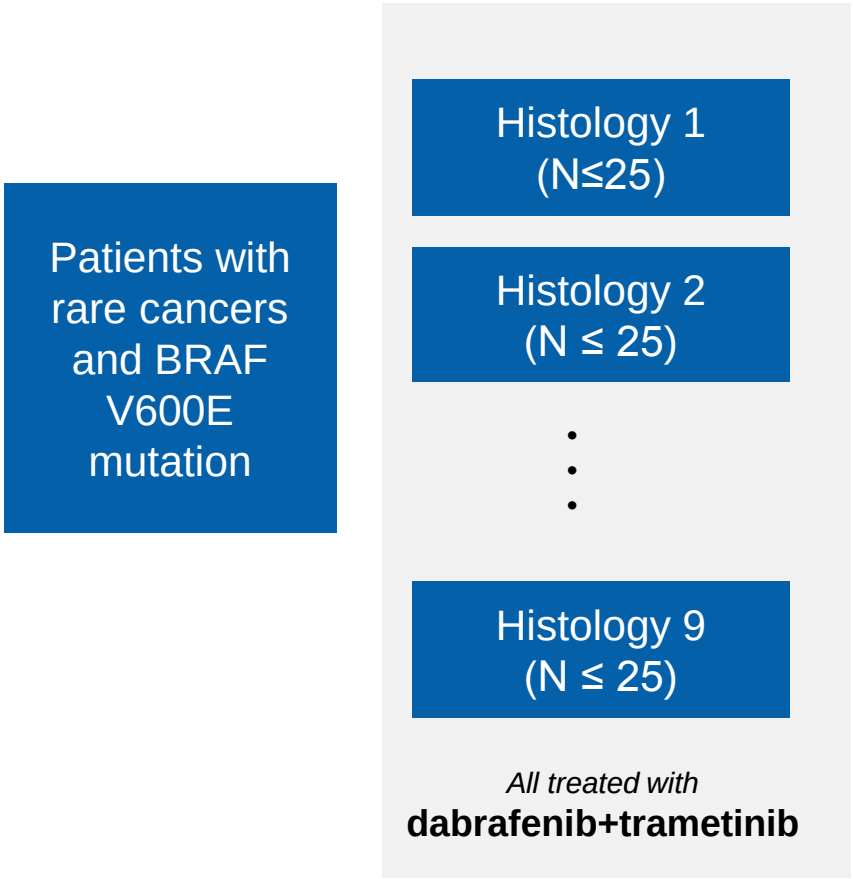
CONS

- 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)

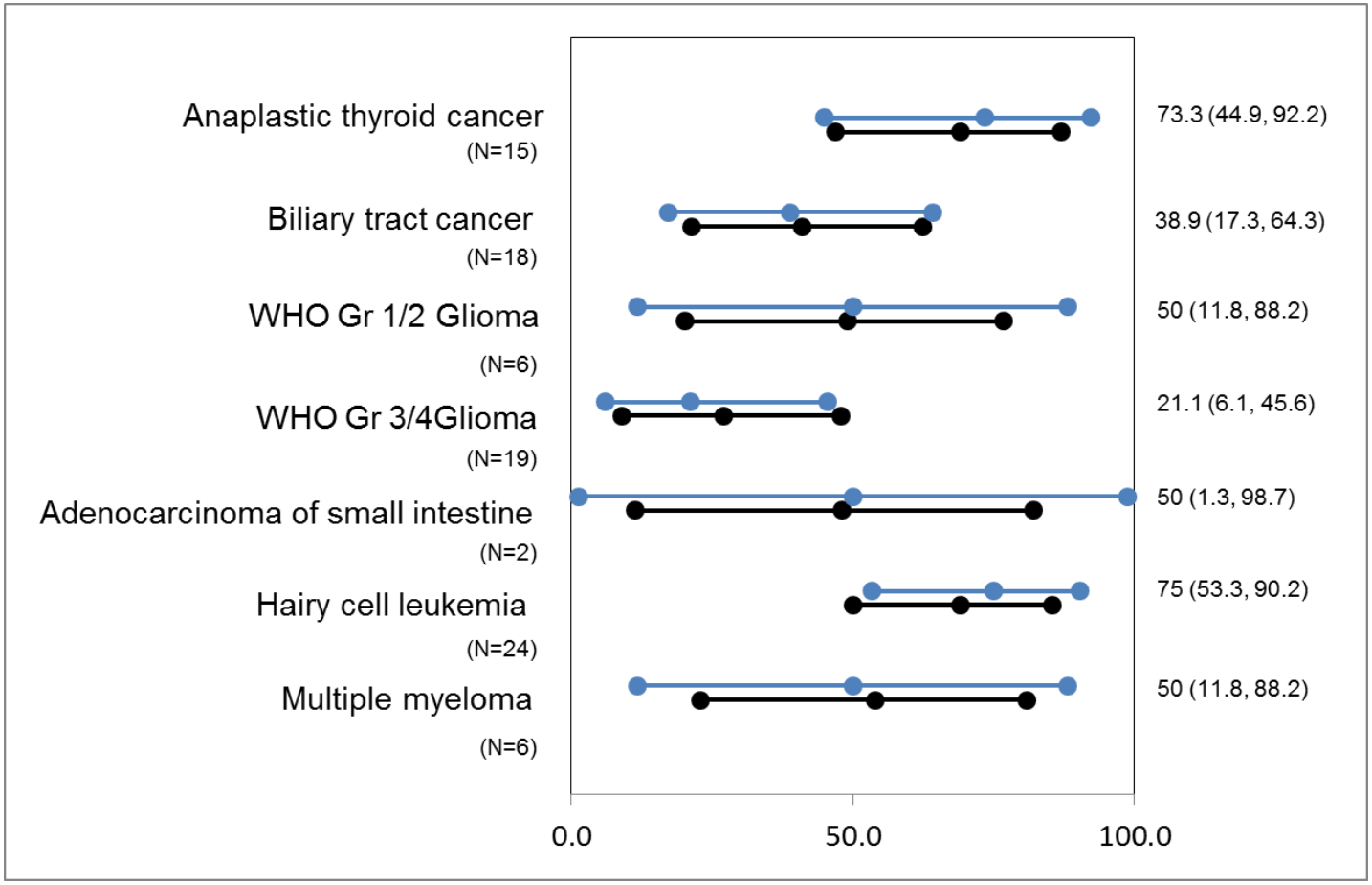
Interim analyses: performed at ~ 12week intervals; enrollment for each cohort may be stopped early for futility or efficacy: decisions based on whether posterior probability that the ORR exceeds its corresponding historical control is sufficiently low or high

* Designed together with Berry Consultants

NOVARTIS

ROAR study – results

Overall response rate



Results presented at ASCO'17.

Observed ORR and 95% exact binomial confidence interval

Estimated ORR and 95% credible interval based on Bayesian Hierarchical model

Biostatistics and Pharmacometrics

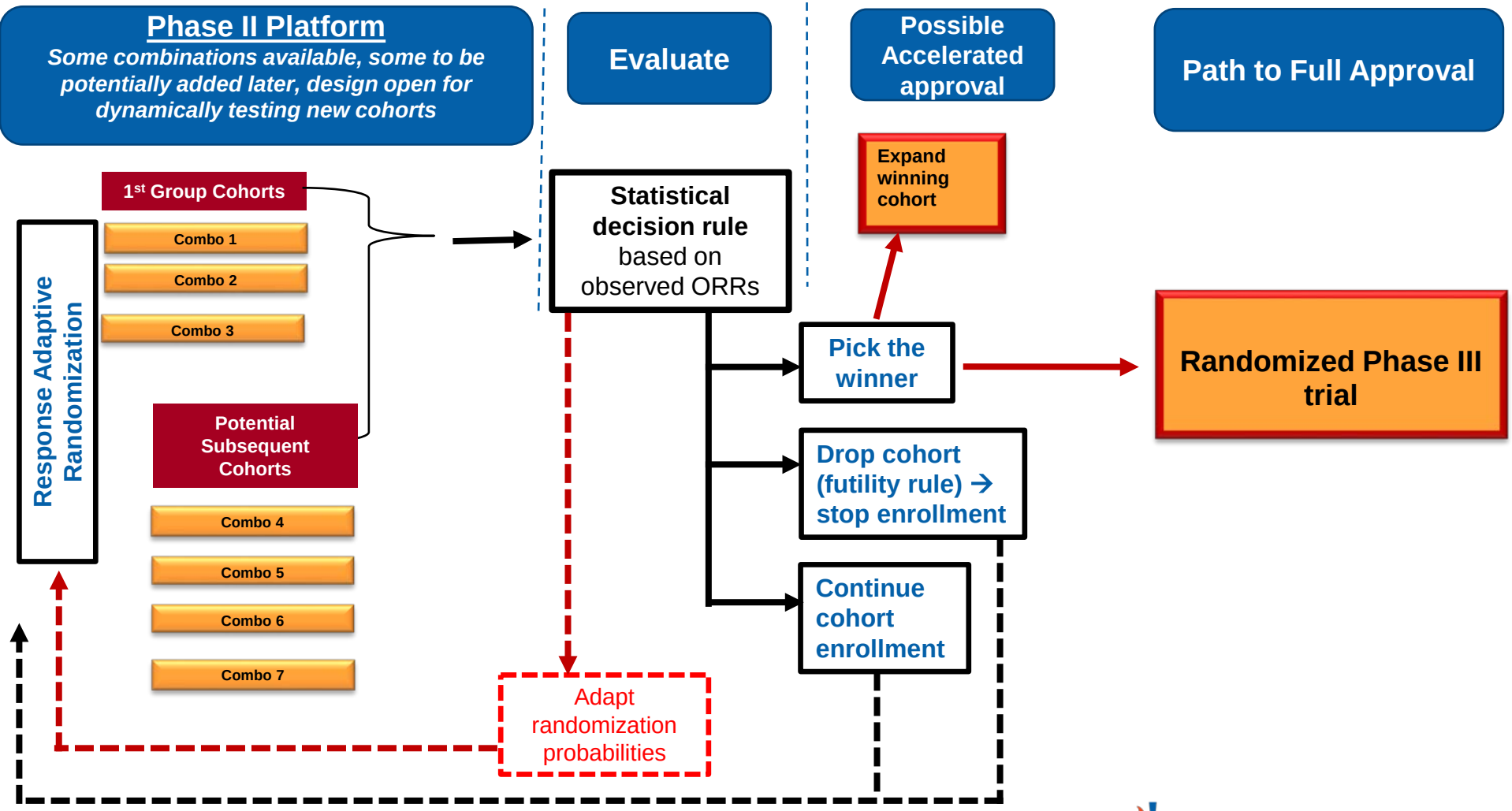
NOVARTIS

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



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, ..., p_k in treatment arms 1, 2, ...k**

- **At completion of batch 1 (total N=n*T)**
  1. **p_i ~ beta($y1_i$+1, $n1_i$-$y1_i$ +1)**
  2. **Decisions**
     - Winner Prob($p_i \geq r_W$|data) $\geq P_W$
     - Futile Prob($p_i \leq r_F$|data) $\geq P_F$
     - Continue otherwise

n = batch size, T 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.

Red numbers are design parameters:

NOVARTIS

Statistical Bayesian model (cont.)

Incorporation of Response Adaptive Randomization (RAR):

1. **For arms which are not clear winners or losers after a batch, calculate**
   - $P_{best_j} = Pr(p_j > p_i, for\ all\ j^{\wedge} = i, continue | response_i)$
   - RAR prob: $P_j = \frac{\left(P_{best_j} * Var(P_j)\right)}{(n_j+1)}$) (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)
   - $P_j = \Pr(Trt\ j\ is\ assigned\ ) = \frac{p_j \left({}^{p_j} \Big/ {}_{\left\{\frac{N_{ji}}{N(i)}\right\}} \right)^{\gamma}}{\sum_{j=1}^{k} p_j \left({}^{p_j} \Big/ {}_{\left\{\frac{N_{ji}}{N(i)}\right\}} \right)^{\gamma}}$, j = 1,....K, N(i) = 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**

NOVARTIS

Design Parameters & Decision Rules Investigated

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

ORR threshold	<x1%	x1-x2%	≥x2%
Decision rule	Prob(ORR<x1%)>p1%		Prob(ORR≥x2%)>p2%
Action	Drop for futility/ stop enrollment	Continue enrollment	Declare 'winner'

Design #	# of arms start/add	n/batch	x1	p1	x2	p2
1	5/2	25	10	80	30	90
2	5/2	25	15	60	20	60
3	5/2	20	15	60	20	60
4	3/1/1	20 with 4 batch total	15	70	20	70
5	4	**8** pts/mth. with 4 batch total	15	70	20	70
6	4/1/1 ("off times")	**8** pts/mth. with 6 batch total	15	70	20	70

Compare against standard design

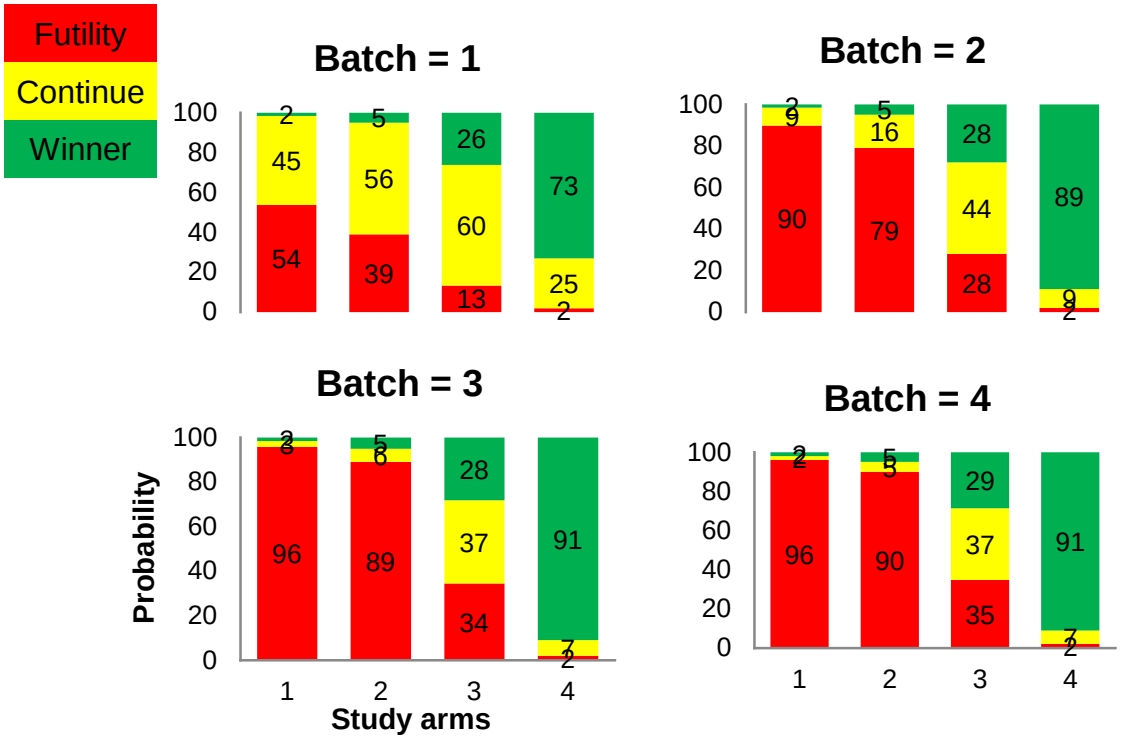
NOVARTIS

Illustration of Design Operating Characteristics

Design 5/Scenario 1: Decision summary and Power

Simulation scenario	4 arms			
	1	2	3	4
True ORR	5%	7%	15%	30%

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



- ‘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

Relatively good OC

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%

NOVARTIS

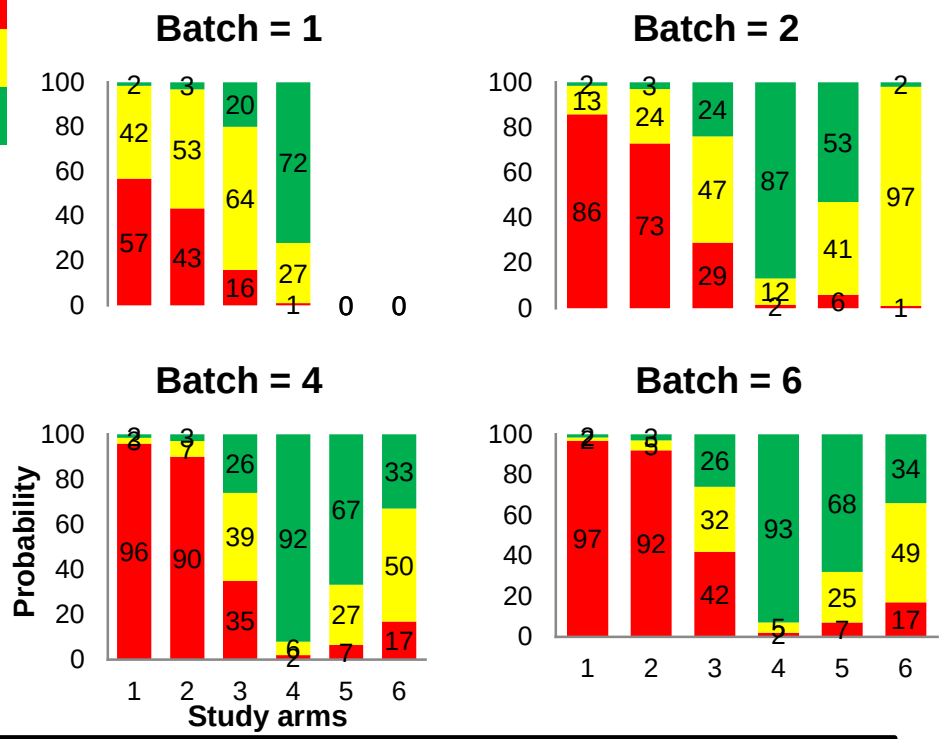
Illustration of Design Operating Characteristics

Design 6/Scenario 1: Decision summary and Power

Simulation scenario	6 arms					
	1	2	3	4	5	6
True ORR	5%	7%	15%	30%	24%	19%

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

Futility
Continue
Winner



- 'Futile' arm 1 (5%):
  57% prob to drop after 1st batch, 86% after 2nd batch; avg N=18
- 'Futile' 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

Relatively good OC

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%

NOVARTIS

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

- Woodcock J and LaVange L (2017). Master Protocols to Study Multiple Therapies, Multiple Diseases, or Both. N Engl J Med 2017; 377:62-70July 6, 2017DOI: 10.1056/NEJMra1510062
- 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
- B. Serville and S Berry (2016) Efficiency of Platform clinical trials: A vision of the future. Clinical Trials.

Back-Up

ROAR study results*

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

[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%

Simulation scenario	4 arms				
Avg. Sample Size	1	2	3	4	Σ
Scenario 1	21	26	30	19	96
Scenario 2	22	26	23	25	96
Scenario 3	23	27	19	16	85
Scenario 4	21	29	28	22	100

‘Futile’ arms

Sample size varies from approximately 21 to 27.

‘Interesting’ arms

Can have sample sizes higher than clear winners and losers; ‘cap’ incorporated to control max.’N’.

‘Winning’ 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

NOVARTIS

Illustration of Sample Sizes

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

Simulation scenario	6 arms					
True ORR	1	2	3	4	5	6
Scenario 1	5%	7%	15%	30%	24%	19%
Scenario 2	5%	7%	5%	7%	4%	20%
Scenario 3	5%	7%	30%	35%	17%	8%
Scenario 4	5%	15%	18%	26%	33%	7%
Scenario 5	11%	16%	21%	26%	19%	12%
Scenario 6	29%	24%	27%	8%	11%	7%

Simulation scenario	6 arms						
Avg. Sample Size	1	2	3	4	5	6	Σ
Scenario 1	18	22	29	17	30	37	153
Scenario 2	18	22	18	23	23	40	144
Scenario 3	18	22	18	15	36	35	144
Scenario 4	18	27	27	21	24	33	150
Scenario 5	25	26	25	20	33	36	165
Scenario 6	18	23	20	23	32	32	148

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

‘Futile’ arms

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

NOVARTIS