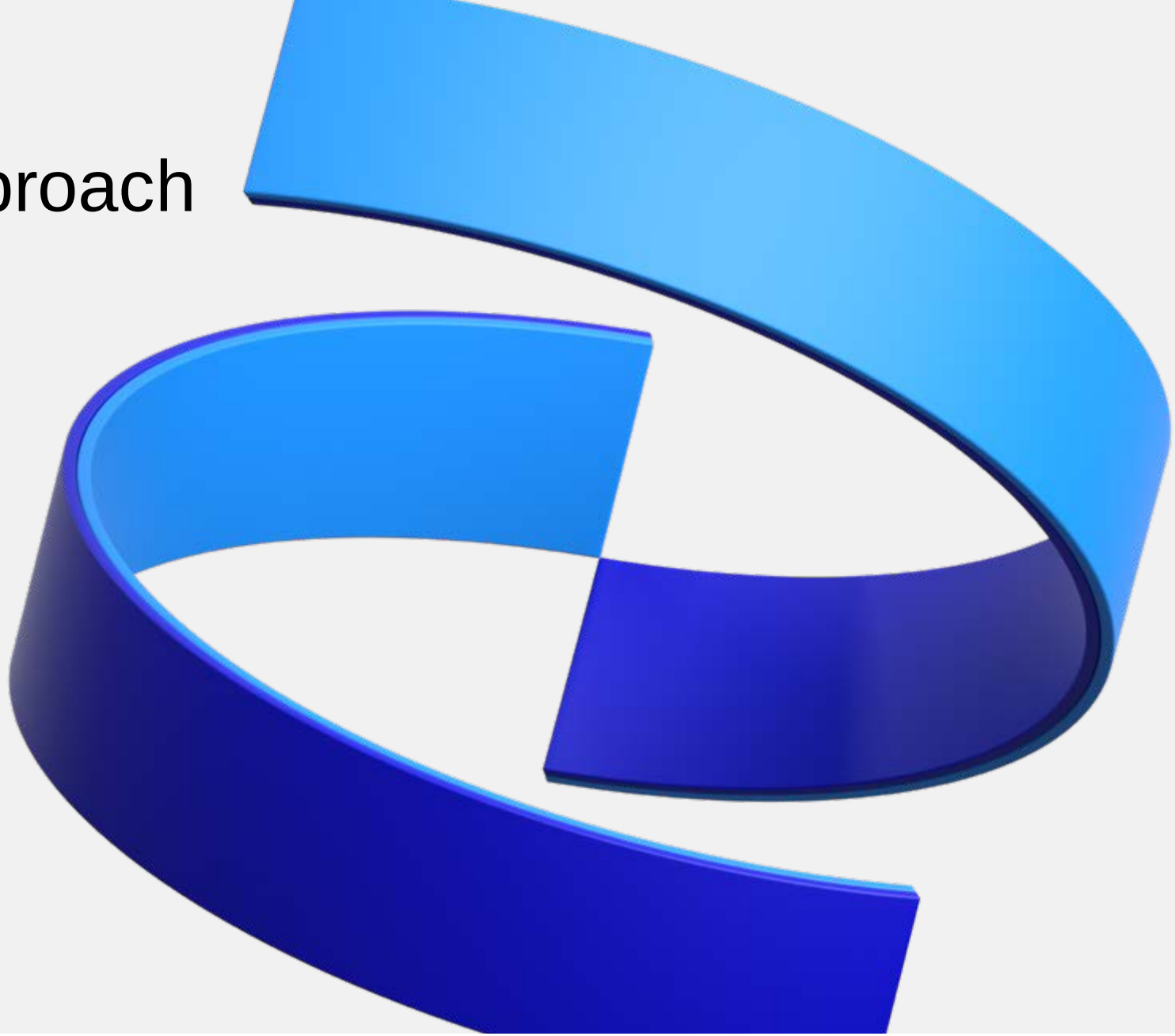
A Robust Bayesian Approach for Basket Trial Design



Satrajit Roychoudhury

Statistical Research and Innovation

Pfizer Inc.

BAYES2022

September 13th, 2022

Recent Advances in Precision Medicine

- Driven by an urgency, an increasing number of agents with potential breakthrough anti-cancer activity have challenged the sequential three-phase paradigm
- Accelerated approval from single arm early phase studies
- Enrichment based on correlative science in early phase trials
- Using trial external information for better decision making
- Innovative and creative thinking is necessary to cope with the need
- Proper risk assessment of innovative approaches is also needed as "*there is no free lunch*"

Master Protocol Design in Clinical Trial

PDUFA VI and 21st Century Cures Act Commitment

- **Complex Innovative Design** program
- FDA pilot program: Sponsors gain increased interaction with agency on design

"Disease characteristic-based" designs (e.g., Biomarker)

- Interaction designs, Enrichment designs, Adaptive enrichment designs, Marker strategy designs
- **Master protocol**
  - Personalized medicine
  - Increased efficiency in drug development when target-drug combinations exist

Master Protocol: Terminology and Definition

Multiple diseases, multiple patient subgroups (biomarker-defined), and/or multiple therapies studied under one, over-arching protocol

Different types of Master protocols

- Basket trial, Umbrella trial, Platform trial
- The terminologies are not straight forward
- Early literature use “basket” and “umbrella” trial inconsistently

More recently, terms became standardized

- Renfro LA and Sargent DJ. *Ann Oncol* 2017
- Woodcock J, LaVange LM. *N Engl J Med* 2017

Recent regulatory guidance

- FDA: Master protocol guidance for Oncology drug and biologics
- FDA: Complex innovative design
- EMA: Clinical evaluation of anti-cancer products

Umbrella and Basket Trial

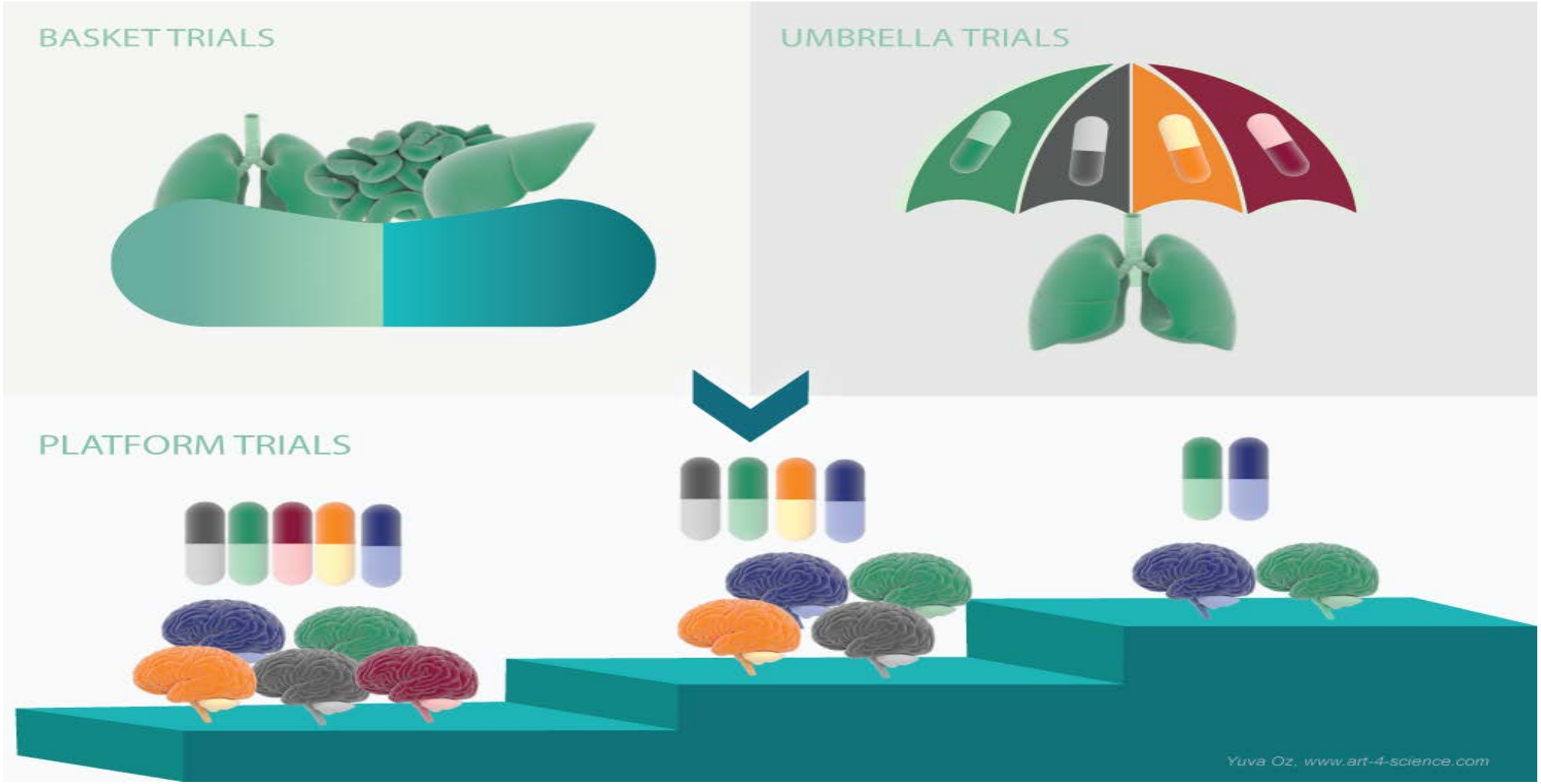
"**Umbrella**" trials

- One tumor type with multiple drugs and predictive biomarkers
- Examples: BATTLE, I-SPY, Lung-MAP

"**Basket**" or "bucket" trials

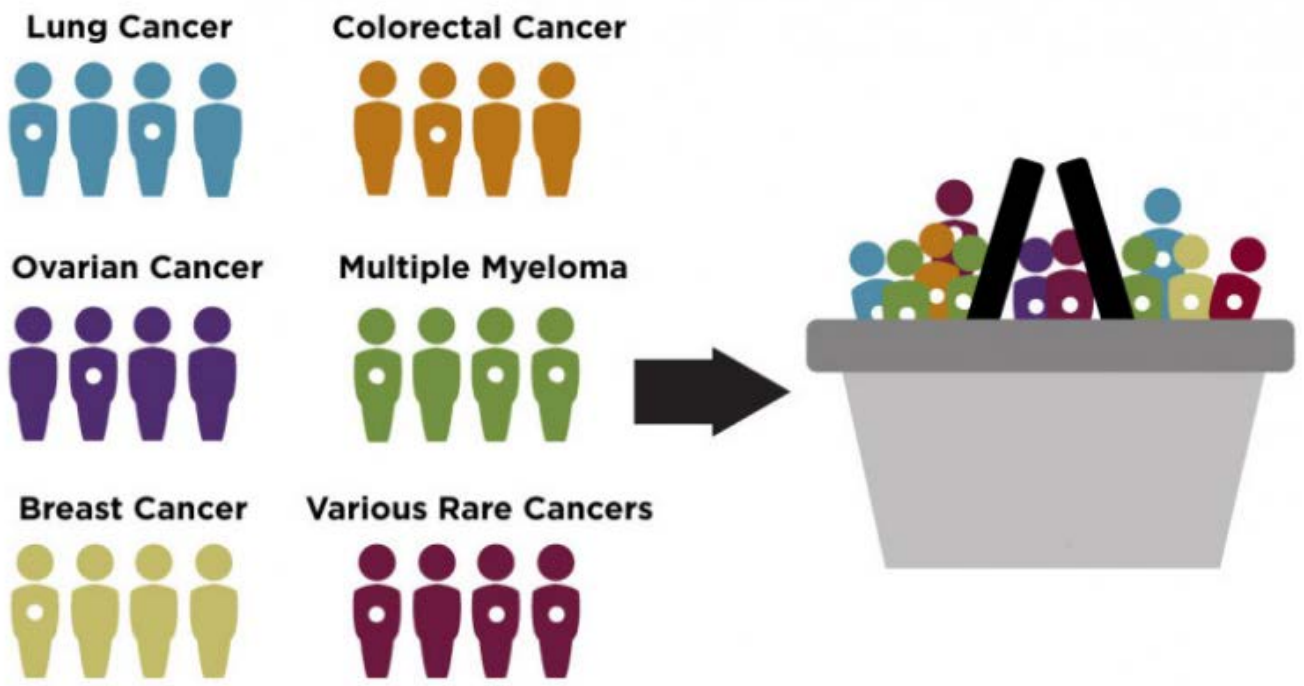
- Multiple tumor types with one drug and a predictive biomarker
- Intends to gain drug approval in multiple tumor types with a common predictive biomarker under the premise that molecular subtype is more fundamental than histology

Master Protocol: Basket, Umbrella and Platform



https://sms-oncology.com/news/blog/the-changing-landscape-of-oncology-clinical-trials-aacr-2017/

Basket Trial: Novel Precision Medicine Trial Design



Basket Trial: Key Features

Typically used in early phase

- Single-arm sub-studies

Preliminary target-treatment hypotheses

- Often uses short term efficacy endpoint

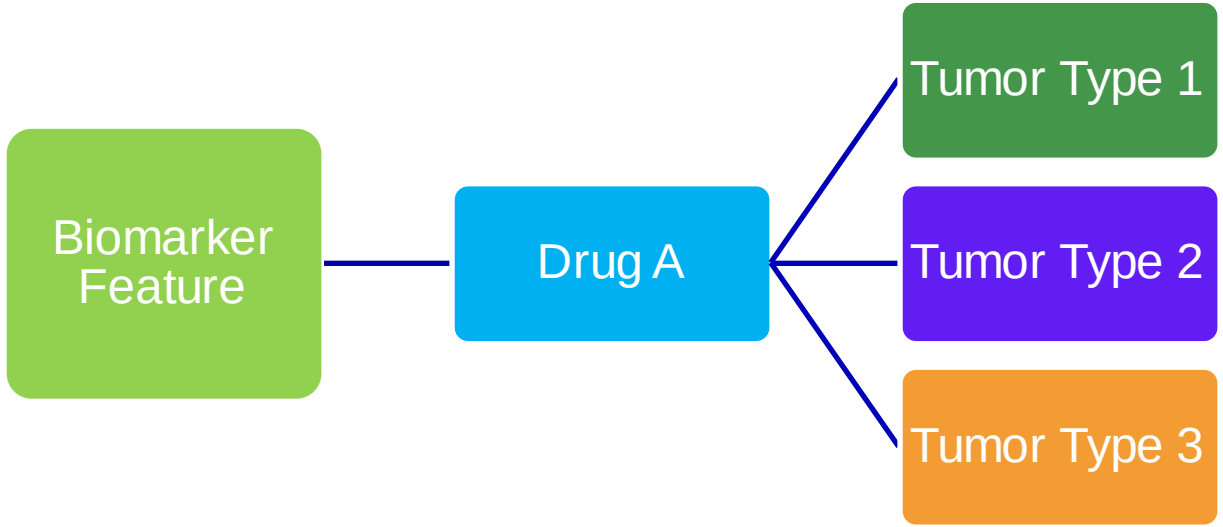
Both single agent and combinations can be tested simultaneously

Target: identify large, unambiguous signals of activity based on molecular features (rather than tumor type)

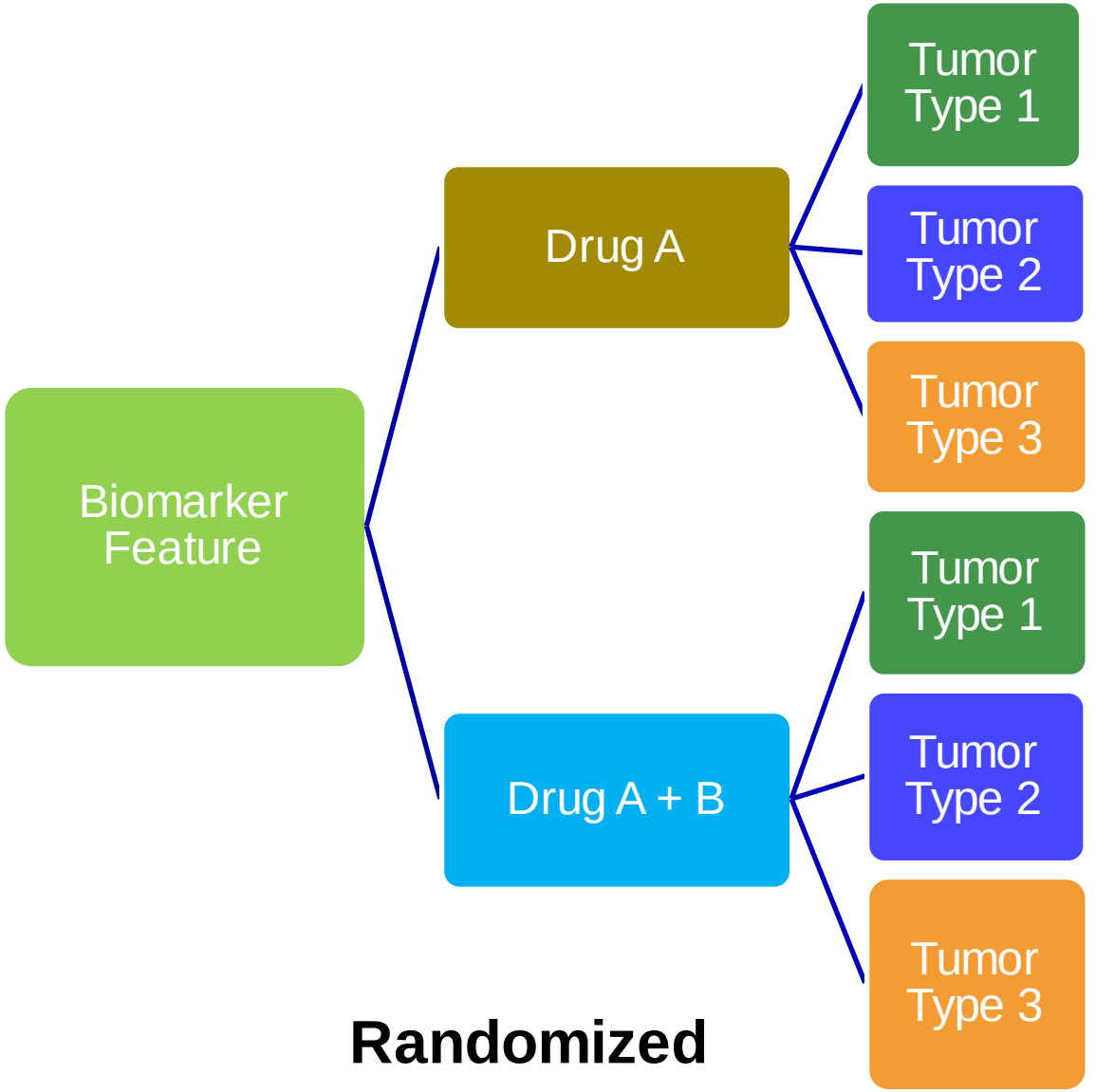
- "Success" within a sub-study may lead to larger confirmatory study

Usually 20-30 patients per "basket" or molecular sub-study

Variations of Basket Trials



Non-randomized



Randomized

Notable Examples

- NCI-Match
- Imatinib B2225 study
- Signature
- Roar
- Pediatric NCI-Match
- AcSe3
- CREATE4

Basket Trial: Key Advantages and Caveats

Advantages

- Increase operational efficiencies due to shared infrastructure
- Better enrollment with rare molecular features across tumor types
- Requires smaller sample size for each tumor type

Caveats

- Key assumption: molecular profiling may be sufficient to replace histological tumor typing
  - Not sufficient evidence
- Prognostic heterogeneity across tumor types
- Methodologically complex with long term endpoint (time to event)

Statistical Aspects for Basket Trial

- Decision making
  - Objective: Primarily PoC; potential accelerated regulatory pathway if overwhelming effect is observed
  - Statistical significance alone may not sufficient to fulfill this objective
  - Formal inclusion of estimation in the decision criteria is necessary
    - Need to plan this at a design stage
    - Dual criteria design: Roychoudhury et. al. 2018, Stein et al. 2011
    - Other aspects: indirect comparison with historical/external control data
- Analysis methods: Frequentist and Bayesian
- Sample size
- Statistical properties (type I error, power, bias, MSE)
- Early stopping for efficacy/futility

Statistical Methodology: Analysis Approaches

Most clinical literature uses two extreme models

- **No pooling:** Separate inference for each tumor type (stratified analysis) - *"Low power for small sample size situations"*
- **Complete pooling:** grouping in the data is irrelevant, i.e. imposing restriction that all tumor type effects are same – "optimistic borrowing"

Better approach: Test and Pool

- Pool only if the tumor types look similar
- Otherwise treat them separate
- Need multiplicity techniques to ensure false negative

Statistical Methodology: Hierarchical Modeling

Bayesian hierarchical modeling is a specific methodology to combine information of different sources

Handles different patterns of heterogeneity

Statistical models often involve multiple parameters grouped together within a multilevel structure

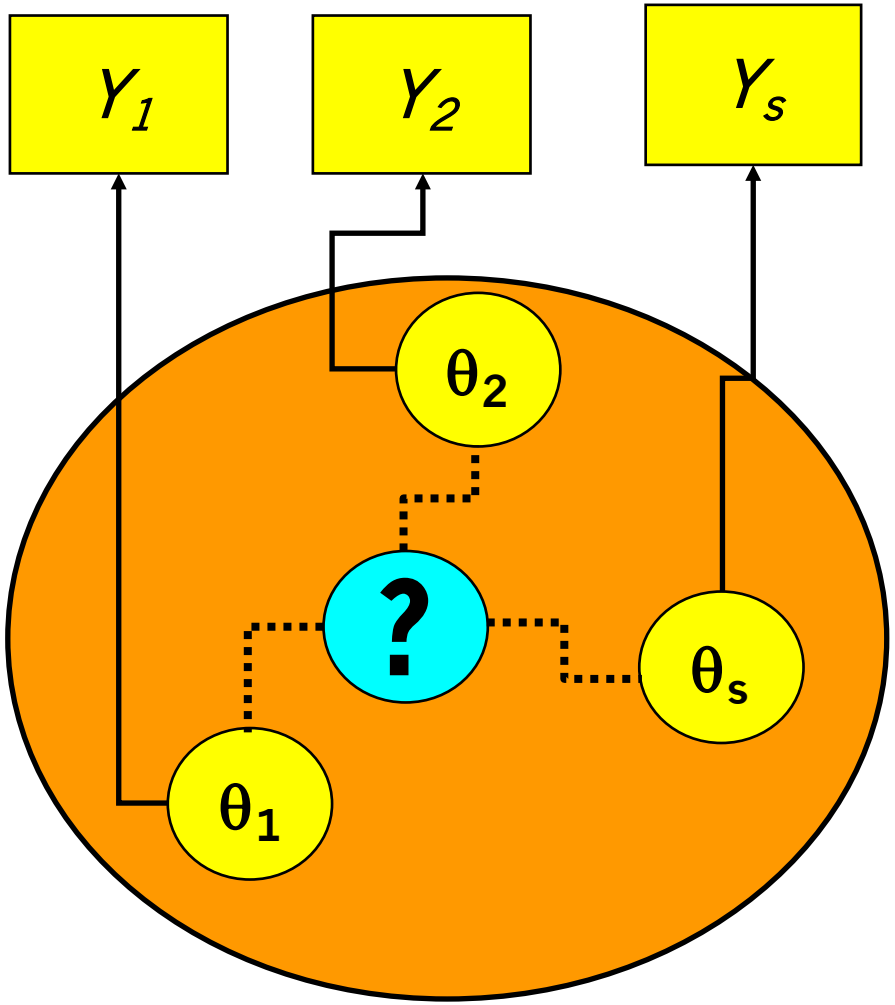
Key assumption: *Exchangeability* or similarity

Lies between these two extreme cases: “No” and “Complete pooling”

This assumption can be relaxed if important disease characteristics are available (e.g. partial exchangeability)

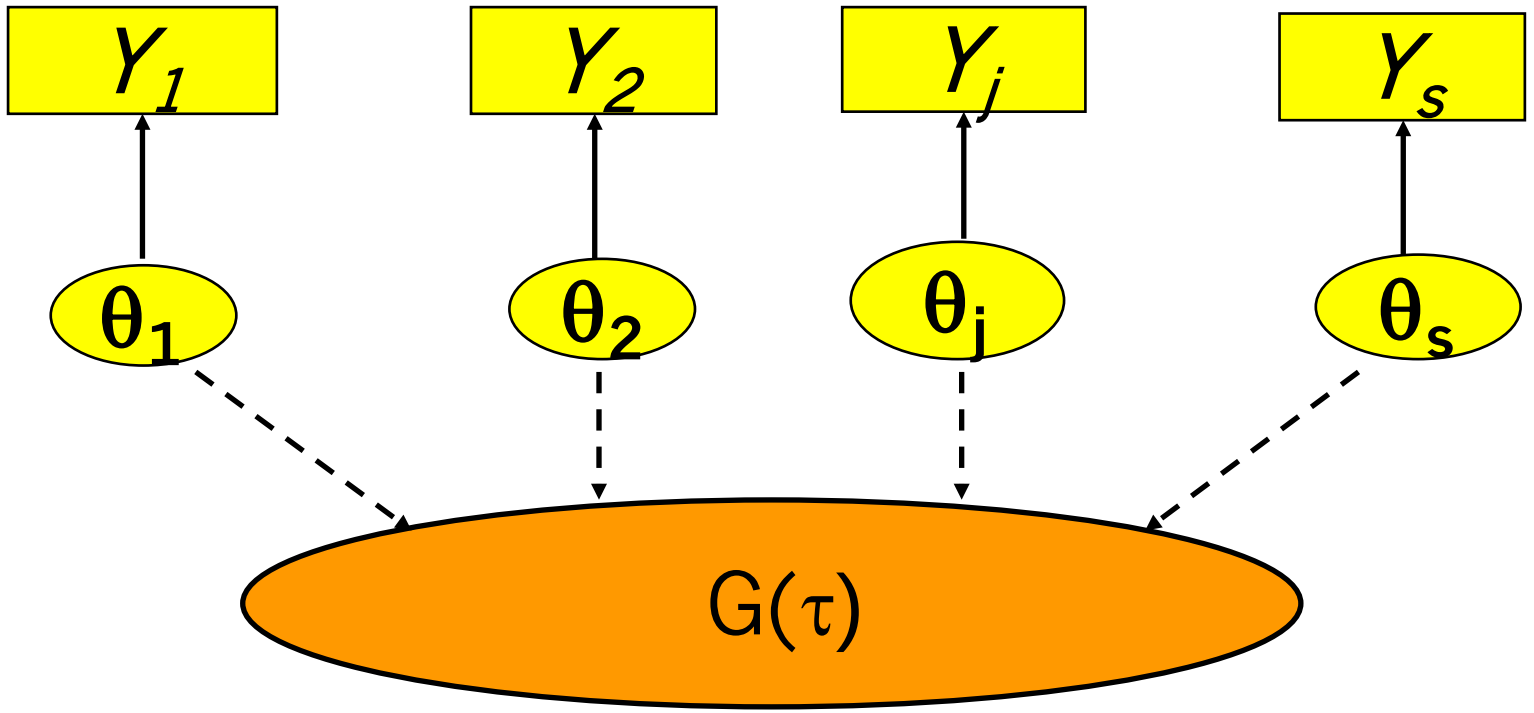
Bayesian “shrinkage” estimates are used to report the treatment effect

Hierarchical Modeling: Key concept



- $Y_1,..,Y_s$: Data from different tumor types
  - Observed ORR
- $\theta_1,\ldots,\theta_s$: tumor-specific parameters of interest
  - True response rate, log-odds
- Questions: Are $\theta_1,\ldots,\theta_s$ related?
  - Exchangeability

Full Exchangeable Hirerchical Model (HM)



- Data (sampling) model $Y_j \mid \theta_j \sim F(\theta_j)$
- Parameter model $\theta_1, \ldots, \theta_s \mid \tau \sim G(\tau)$
- Shrinkage estimates are pulled towards a common mean (too restrictive?)

Prior distributions for τ

- Since the number of arms is often small: prior matter
- Recommendations (Spiegelhalter 2004, Gelman 2006)
  - Use priors that put most of their probability mass on plausible values

Endpoint	Very conservative τ prior	Conservative τ prior	μ prior (Unit information prior)
Binary	HN(0, 1^2)	HN(0, 0.5^2)	N(0, 2^2)
Normal	HN(0,$(\sigma/2)^2$)	HN(0,$(\sigma/4)^2$)	N(μ_0, σ^2)

Effective Sample Size (ESS)

- Expected Local-information-Ratio (ELIR)
  - Information-based
  - ESS_{ELIR} defined as the expected (under the posterior) ratio of posterior information $i(p(\theta|y))$ to Fisher information $i_F(\theta|y)$

$$n_{**} = E_\theta\{\frac{i(p(\theta|y))}{\mathrm{i_F}(\theta|y)}\}$$

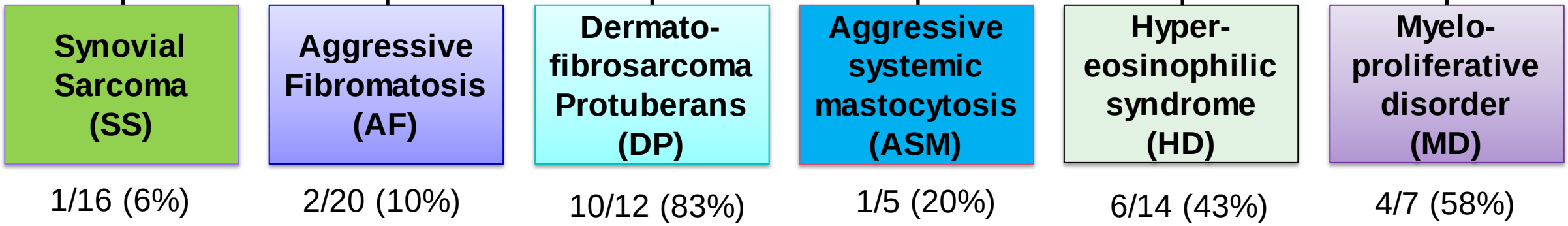
  - Ensure predictive consistency: posterior ESS must be the sum of the sample size and amount borrowed (Neuenschwander et. al. 2019)

An Example: Imatinib B2225 study

186 subjects with 40 different malignancies with known genomic MOA of imatinib target kinases

KIT, PDGFRA, or PDGFRB

Primary endpoint: ORR



Blumenthal. Innovative trial designs to accelerate the availability of highly effective anti-cancer therapies: an FDA perspective, AACR 2014

HM for the B2225 study

- **Data:** n_j = Number of patients and r_j = Number of responder for strata j

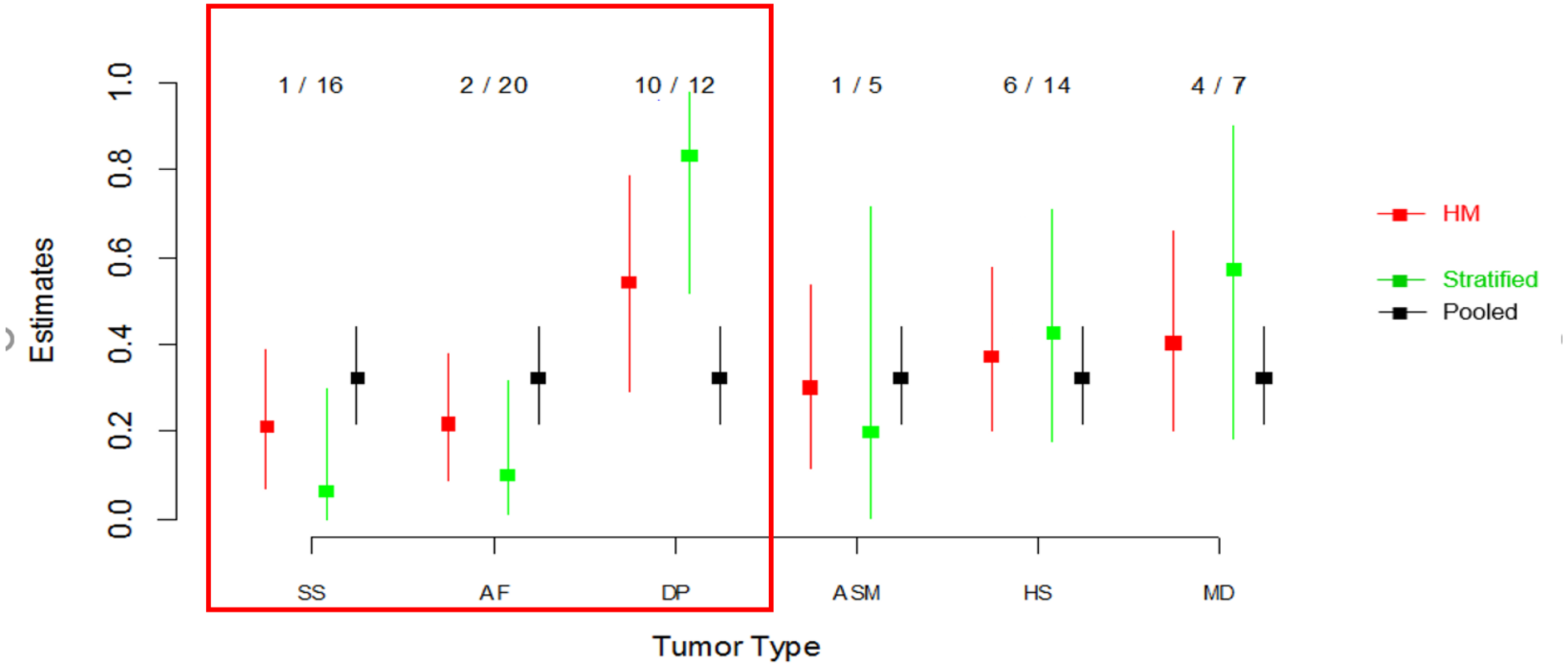
- **Likelihood/sampling model:** $r_j \sim Bin(n_j, \pi_j)$

- **Model:** $\theta_j = \log(\pi_j / 1- \pi_j)$

For each stratum j two possibilities are considered: $\theta_j \sim N(\mu,\tau^2)$

- Priors: $\mu \sim N(0, 10^2)$; $\tau \sim HN(1)$

Analysis of B2225 study Using HM



Robust Hierarchical Model (Robust HM)

Assumption that molecular profiling may be sufficient to replace histological tumor typing

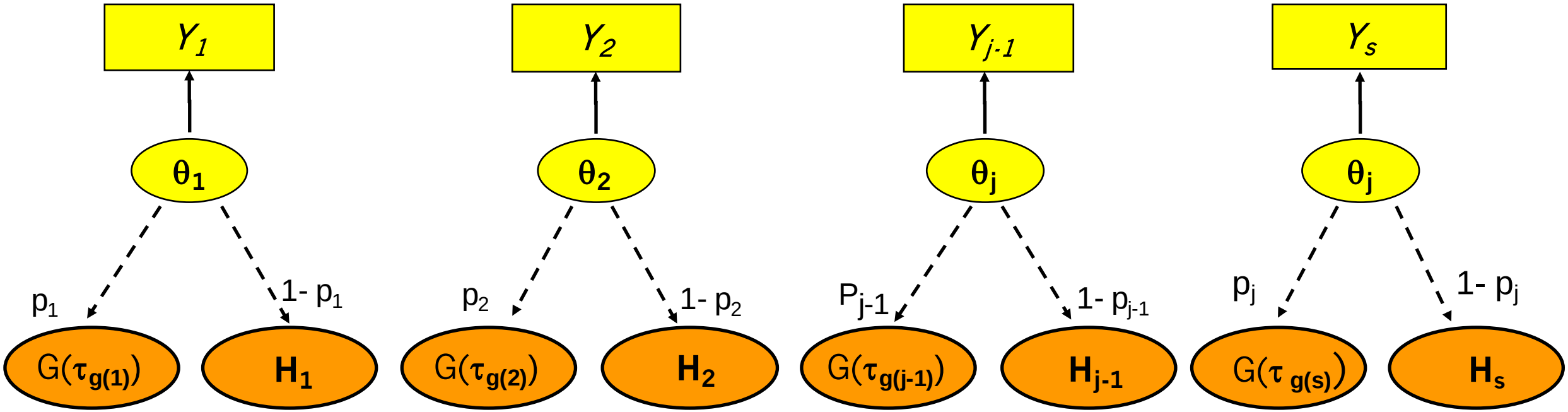
- This may not be the case: V600E BRAF-mutant melanoma vs colon
- Variability among tumor types due to small sample

Full exchangeability of strata parameters may be too restrictive

We propose Tailored exchangeability or mixture model (**EXNEX**)

- Borrowing information across similar strata, while avoiding too optimistic borrowing for extreme strata
- Reduce misclassification error in mixed situation: reduce the “shrinkage effect”
- Neuenschwander, Roychoudhury, and Schmidli 2016, Neuenschwander, Wandel, Roychoudhury, and Bailey 2016
- ASCEND 10 trial: Kiss et 2016

Robust BHM for Dynamic Borrowing Across Tumor Types



- Robustification: $\theta_j \sim p_j\, G(\tau_{g(j)}) + (1- p_j)\, H_j : g(j) \in \{1,...,G\}, j = 1,...s$
- Allows for nonexchangeable parameters to add robustness

Robust BHM as Model Averaging

- Mixture components can be thought of as representing 2 competing models about treatment effect
  - π_1: ***exchangeable*** *probability aprioi*
  - π_2: **non-exchangeable**
- ***w*** represents apriori probability of each model
- Posterior probability update by Bayes theorem with the data (y)
- $w^* = P(\pi_1|y) = \frac{f(y|\pi_1)P(\pi_1)}{f(y|\pi_1)P(\pi_1)+f(y|\pi_2)P(\pi_2)}$;

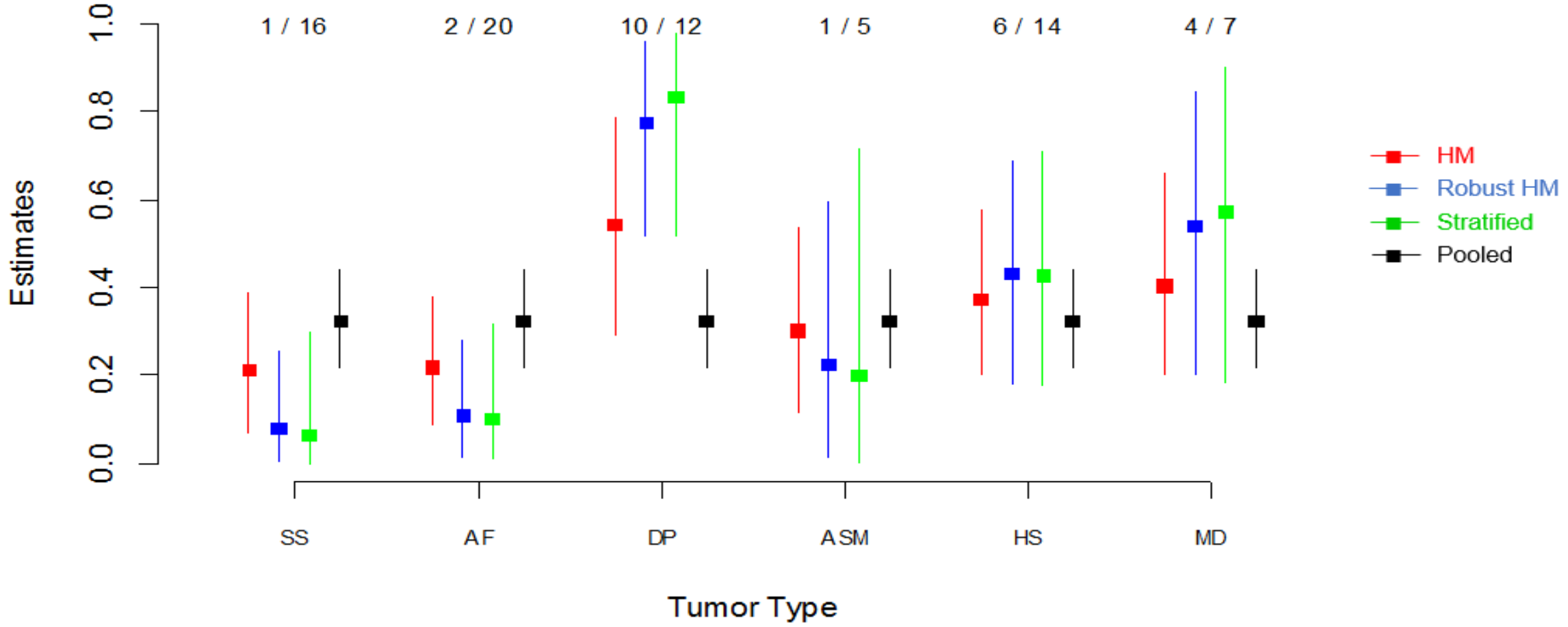
  $f(y|\pi_1)$: likelihood under model π_1
- Robust BHM adopts model averaging techniques

Robust HM for The B2225 study

- **Data:** n_j = Number of patients and r_j = Number of responder for strata j
- **Likelihood/sampling model:** $r_j \sim Bin(n_j, \pi_j)$
- **Model:** $\theta_j = \log(\pi_j / 1-\pi_j)$

  For each stratum j two possibilities are considered:
  - With probability p_j : $\theta_j \sim N(\mu, \tau^2)$
  - With probability $1-p_j$: $\theta_j \sim N(\mathbf{m_J}, \mathbf{v_J})$
  - $p_j = 0$ => HM
  - For this example we assume $p_j = 0.5$

Analysis of B2225 study Using Robust HM



Dirichlet Process Mixture Model

- Class of nonparametric model for clustering
- Population of θ's come from flexible distribution of "clusters"
- Hierarchical models over the cohorts within each cluster to determine the appropriate extent of borrowing between cohorts
- Does not allow borrowing across clusters: minimal borrowing across dissimilar subgroups
- Entirely data driven: number of cluster is not pre-specified
- Reduce misclassification errors relative to the basic hierarchical models in mixed situation
- Signature trial: Slosberg 2018

Simulation Study

A simulation study to compare different methods:

- No-pooling or stratified analysis (STRAT)
- Full exchangeability model (EX)
- Robust exchangeable models
  - 10% chance of apriori non-exchangeable (EXNEX90)
  - 50% chance of apriori non-exchangeable (EXNEX50)
- Dirichlet process mixture model (DPM)

Simulation design

- **4** strata with **n= 20**: Binary outcome
- Declare success: **P(Response rate > 15% | data) > 0.95**
- A response rate of 45% or higher is of clinical interest (target)

Simulation Scenarios

Scenario	T1	T2	T3	T4
1	15%	15%	15%	15%
2	45%	15%	15%	15%
3	45%	45%	15%	15%
4	45%	45%	45%	15%
5	45%	45%	45%	45%

Type I Error: Familywise Error Rate

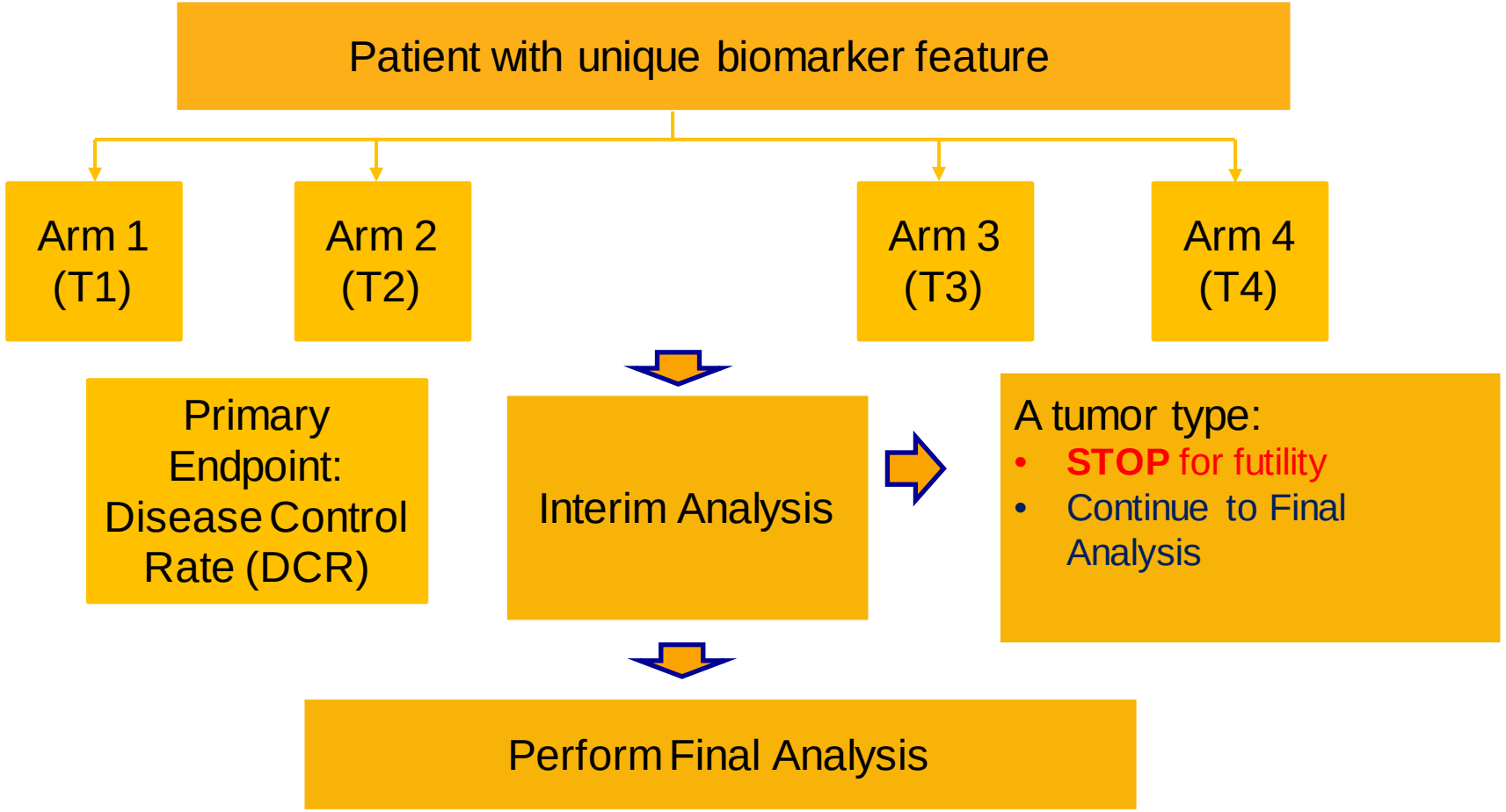
Scenarios	STRAT	EX	EXNEX90	EXNEX50	DPM
1	0.11	0.04	0.07	0.11	0.07
2	0.09	0.16	0.14	0.13	0.13
3	0.05	0.21	0.21	0.11	0.15
4	0.02	0.35	0.16	0.04	0.15

Power Comparison

Scenarios	STRAT	EX	EXNEX90	EXNEX50	DPM
2	0.74	0.75	0.80	0.84	0.77
3	0.70	0.83	0.76	0.74	0.76
4	0.68	0.93	0.94	0.85	0.88
5	0.59	0.98	0.98	0.94	0.97

EXNEX50 seems to be a good candidate based on operating characteristics

Example: Design with Rare Tumor Types



Sample Size and Decision Rules

Tumor Type	Sample Size	Interim Analysis	C1	C2
T1	30	10	≤40%	≥50%
T2	30	10	≤20%	≥30%
T3	30	10	≤10%	≥20%
T4	30	10	≤10%	≥20%

Sample size is **30** per tumor size

Declare **trial success**

- P(DCR ≤ C1 | data) < **10%**
- Posterior mean > **C2**

STOP for futility

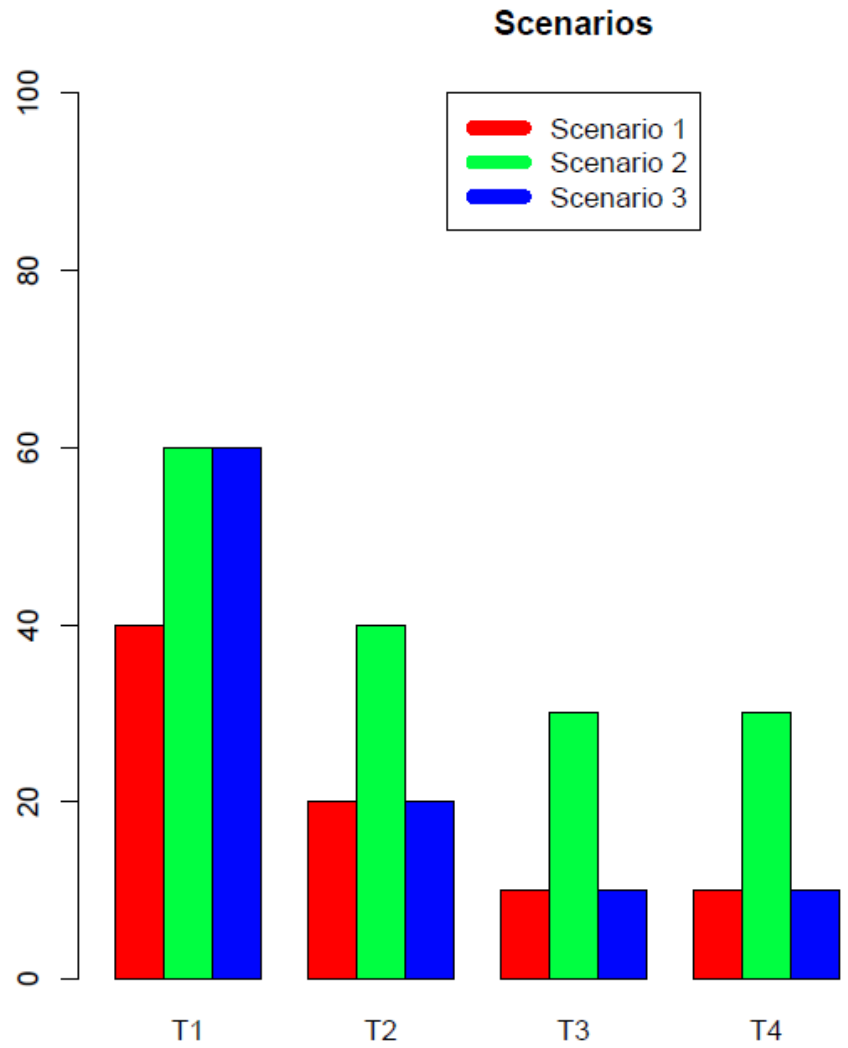
- P(DCR ≥ C2 | data) < **20%**

For any analysis all accumulated data will be used

All analysis are based on an **EXNEX model**

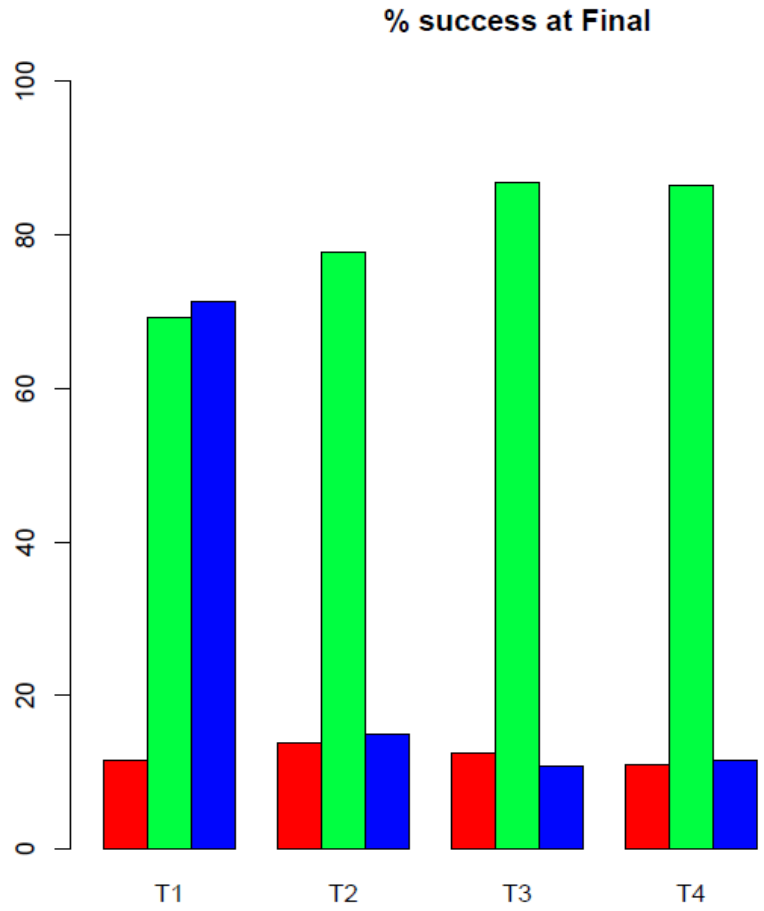
- 50% chance of apriori non-exchangeable

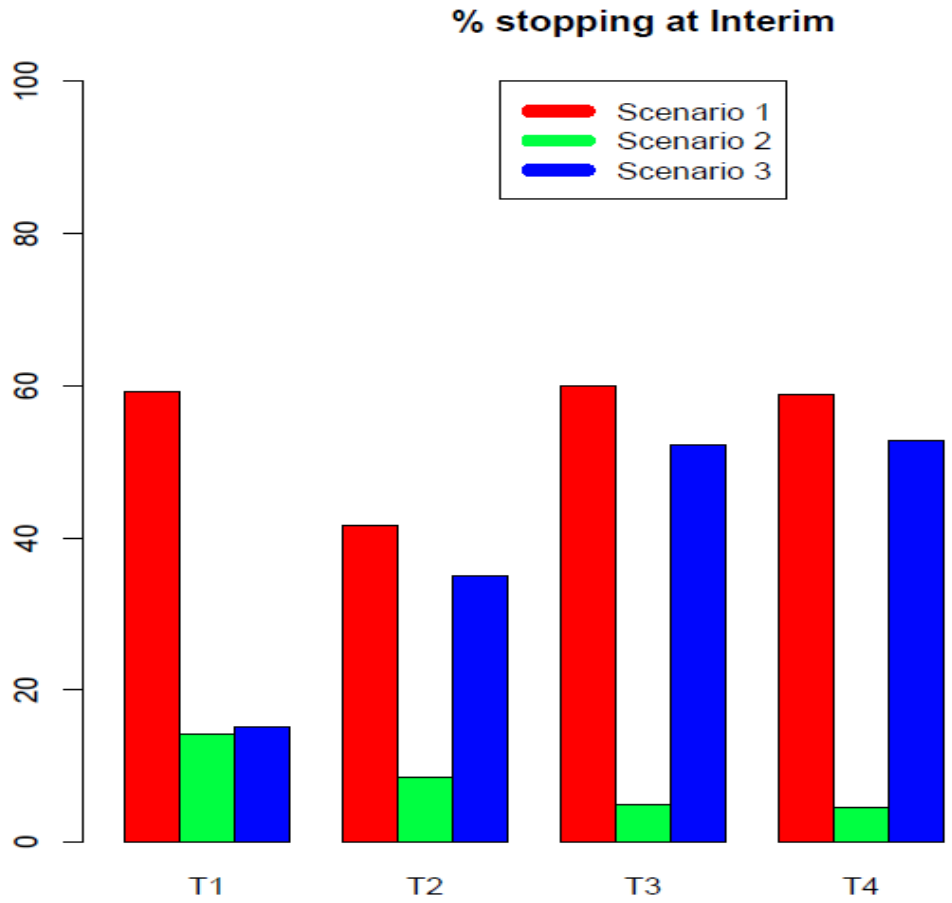
Simulation Scenarios



- Evaluated under 3 different scenarios
  - Scenario 1 : **not clinically meaningful** for any arm.
  - Scenario 2 : **clinically meaningful** for all arms
  - Scenario 3 : clinically meaningful for T1
- Accrual rates are in proportion 2:1:2:2 for the 4 arms (T1, T2, T3 and T4)

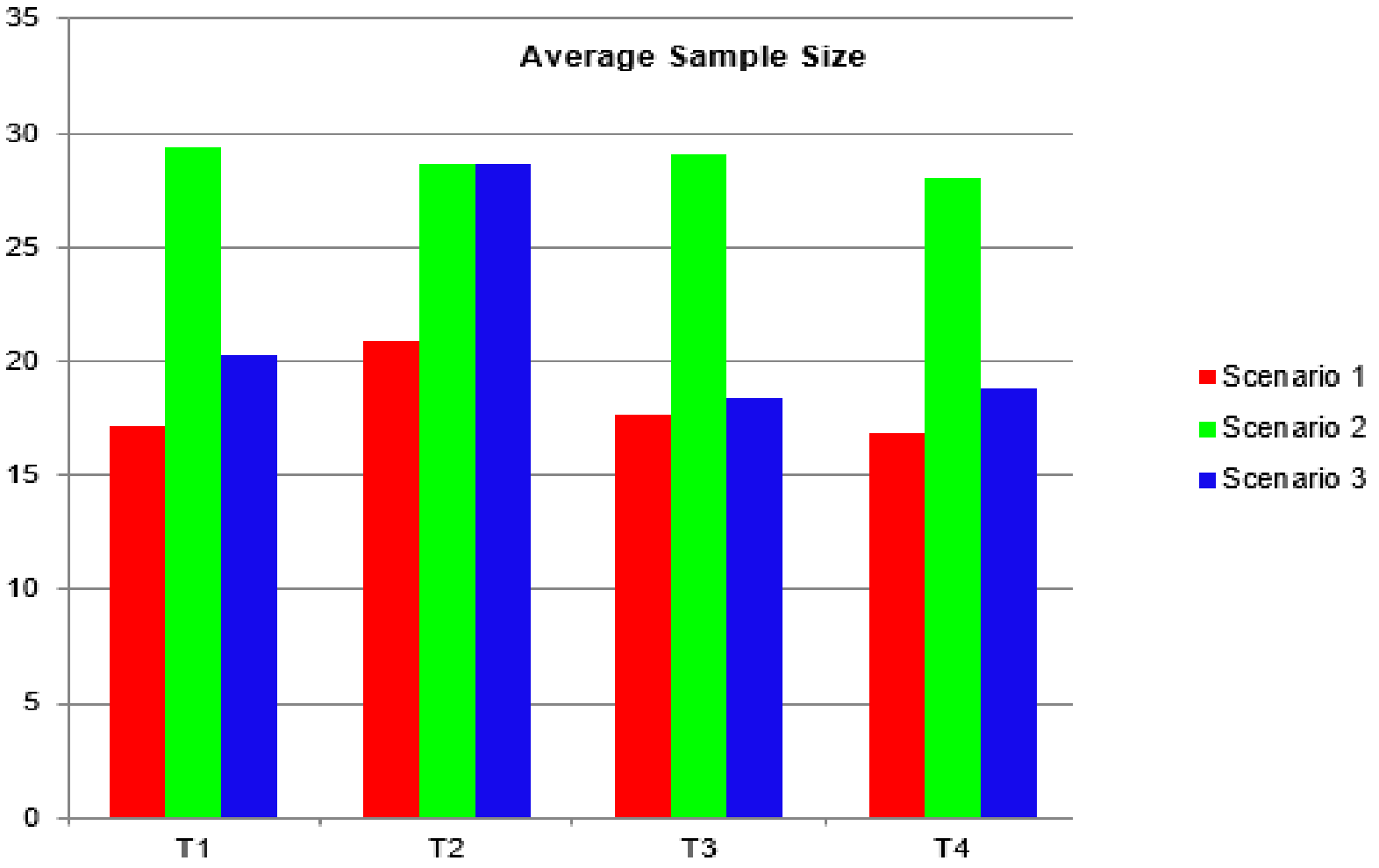
Simulation Results: Decision at Interim and Final





- Reasonable stopping probabilities for different scenario
- Acceptable Type-I error control (<=11% max) and power

Simulation Results: Average Trial Sample Size



Summary

Hierarchical borrowing allows for extra power with multiple subgroups

Possible caveats

- Lower power in “nugget” scenarios
- Inflation of type-I error in “mixed” scenarios

Due to robust nature of EXNEX dynamic borrowing of information between arms

- Implies more borrowing when the arms are consistent and less borrowing when the arms are different
- Significantly mitigates the problems mentioned above
- **A good candidate for Basket trial design**

Other possibilities include: random partitioning model (Leon-Novelo et. al. 2013), multi-source exchangeability model (Kaizer et al. 2018)

Acknowledgement

1. Beat Neuenschwander, Novartis Pharma AG

2. Xin Zhang, Pfizer Inc

References:

1. Neuenschwander, Roychoudhury, and Schmidli (2016) On the Use of Co-Data in Clinical Trials, Statistics in Biopharmaceutical Research, 8:3, 345-354.

2. Neuenschwander, Wandel, Roychoudhury, and Bailey (2016) Robust exchangeability designs for early phase clinical trials with multiple strata. Pharmaceutical Statist., 15: 123– 134.

3. Roychoudhury, Scheuer, and Neuenschwander (2018). Beyond p-values: A phase II dual-criterion design with statistical significance and clinical relevance. Clinical trials 15 (5): 452-461 .

Thank You