



MRC  
Biostatistics  
Unit



UNIVERSITY OF  
CAMBRIDGE

---

# Adaptive Methods in Clinical Research

## *Lecture 4: Master protocols*

Thomas Jaki

MRC Biostatistics Unit

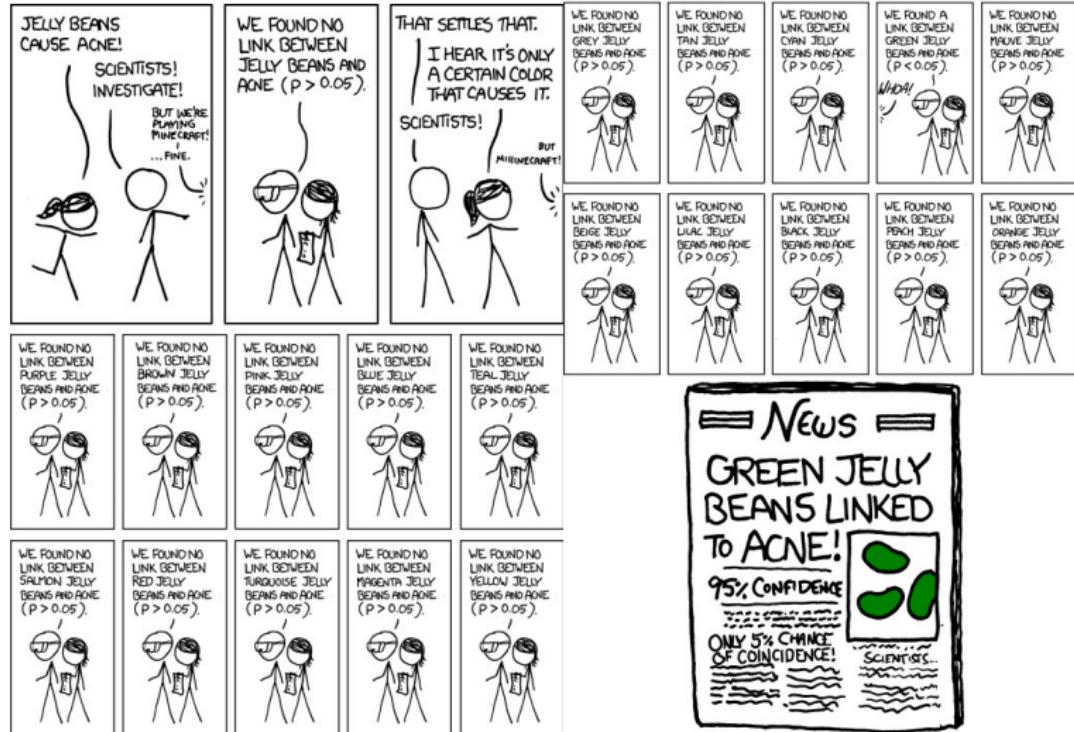
18 September 2024

# Master protocols

Master protocols are a *new* type of study that seeks to answer multiple questions within a single study

- Platform trials
- Basket trials
- Umbrella trials

# Multiple testing



<https://xkcd.com/882/>

# Multiple testing

## **Family-wise error rate (FWER)**

Weak control

$$\mathbb{P}_\theta(\text{Reject at least one true null hypotheses} | \mathcal{H}_0) \leq \nu.$$

Strong control

$$\max \mathbb{P}_\theta(\text{Reject at least one true null hypotheses}) \leq \nu.$$

## **False discovery rate (FDR)**

Controls proportion of false discoveries among the discoveries

# When to adjust for multiplicity

Parker & Weir (2020)

- Other treatment arms irrelevant for making decision about a particular treatment  
⇒ **Do not adjust**
- When testing different subgroups, doses, ...  
⇒ **Consider adjusting**

# When to adjust for multiplicity

## Parker & Weir (2020)

- Other treatment arms irrelevant for making decision about a particular treatment  
⇒ **Do not adjust**
- When testing different subgroups, doses, ...  
⇒ **Consider adjusting**

## Bretz & König (2020)

- Not all null hypothesis are simultaneously true (weak control)
- Multiple hypotheses are often tested for a single submission (e.g. Basket trial)  
⇒ **Adjust in confirmatory setting**

# When to adjust for multiplicity

## Parker & Weir (2020)

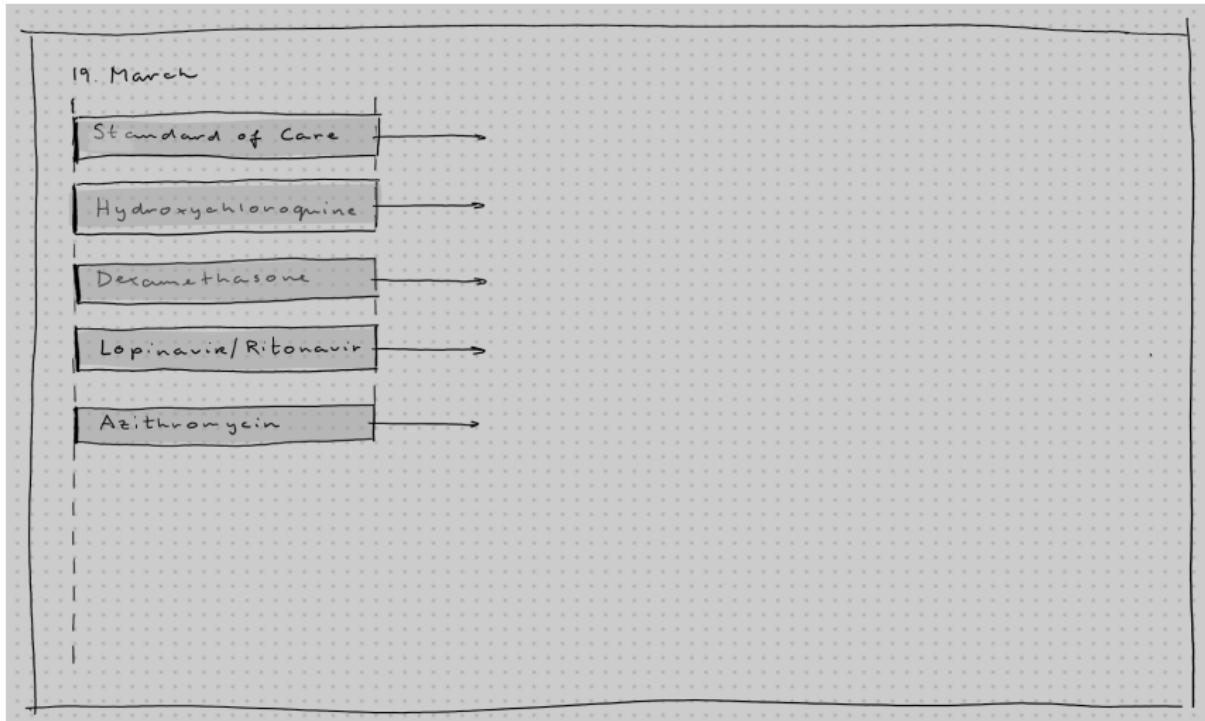
- Other treatment arms irrelevant for making decision about a particular treatment  
⇒ **Do not adjust**
- When testing different subgroups, doses, ...  
⇒ **Consider adjusting**

## Bretz & König (2020)

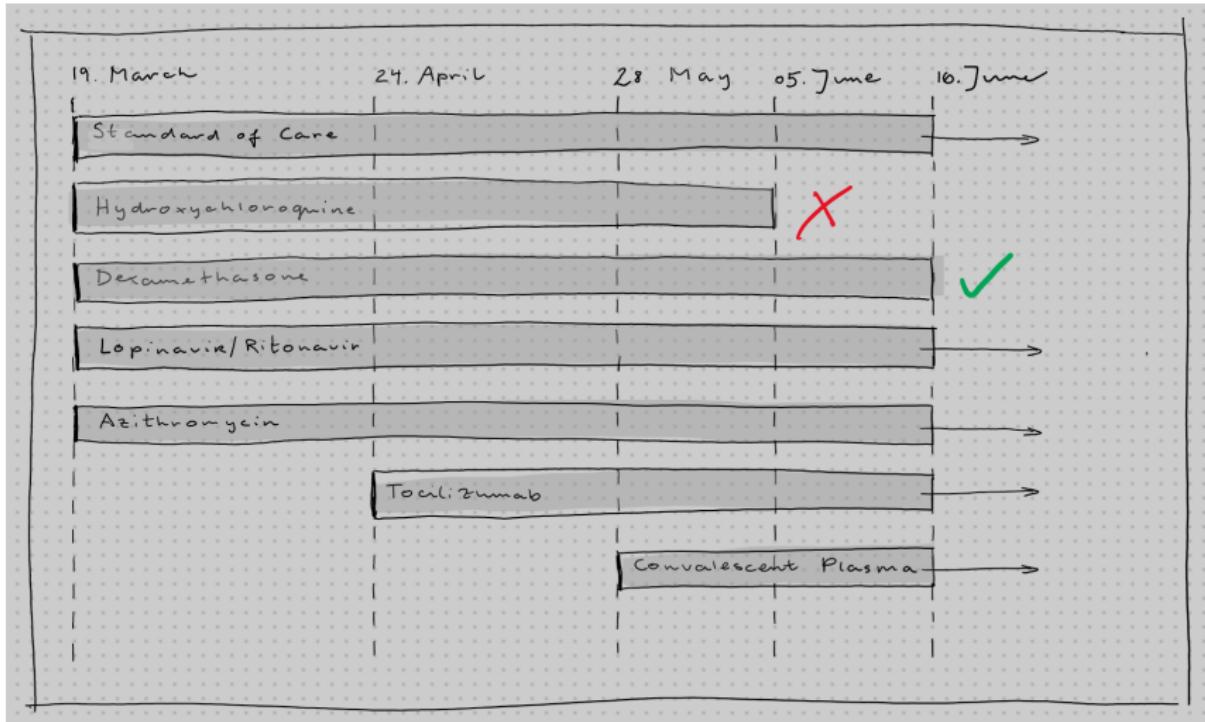
- Not all null hypothesis are simultaneously true (weak control)
- Multiple hypotheses are often tested for a single submission (e.g. Basket trial)  
⇒ **Adjust in confirmatory setting**

**Need to understand risk of false decisions**

# A platform trial (RECOVERY)



# RECOVERY



# What you probably didn't know

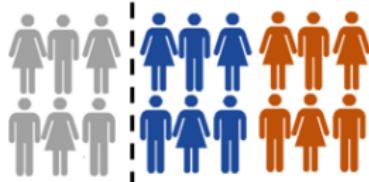
- Initial design of RECOVERY set to control FWER using methods discussed in previous lecture
- At the time not clear how to do so when adding arms

# Sampling scheme (2 stage)

In stage 1 we recruit patients to all of the treatments



We then conduct an interim analysis to choose which treatments to continue recruiting from



## At interim analysis $j$

- if  $Z_{k,j} < l_j$ : treatment  $k$  is dropped from trial.
- if  $Z_{k,j} > u_j$ : can reject  $H_k$  and stop trial.

# Family-wise error rate (Magirr et al, 2012)

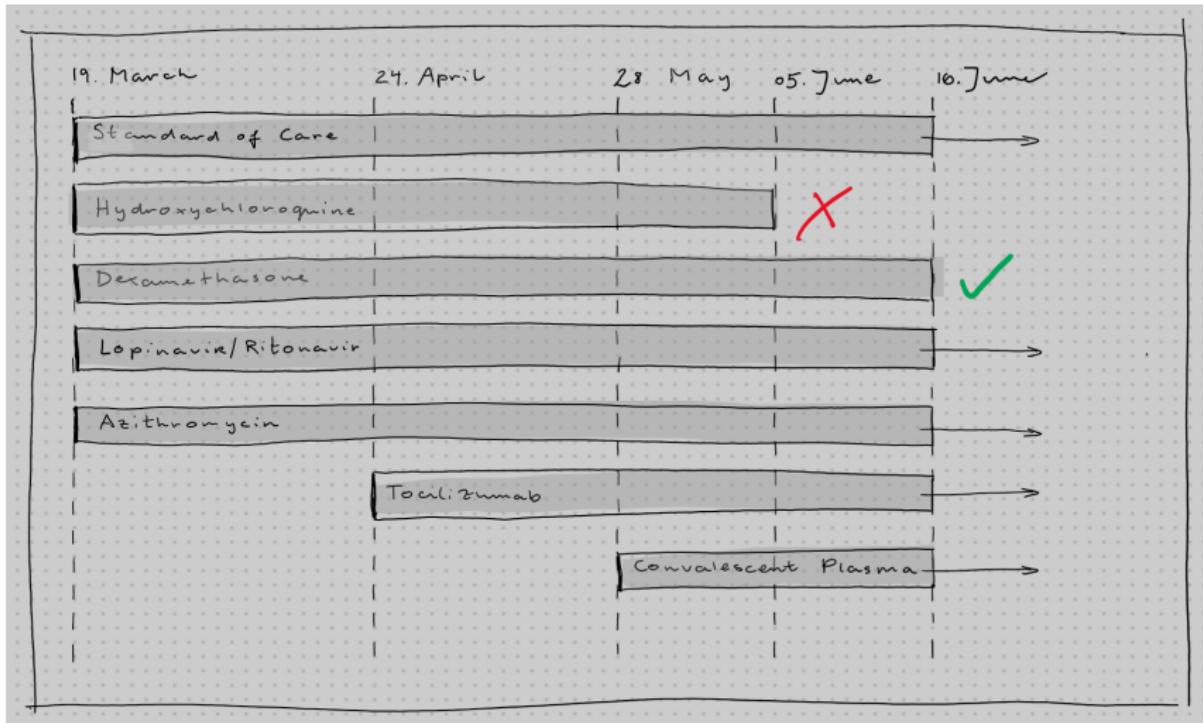
Issue:  $J$ -stage trial with  $K$  arms  $\Rightarrow$  up to  $K * J$  tests.

Fact: Strong control of FWER  $\Leftrightarrow$  Weak control of FWER.

Problem: Test statistics are correlated.

Solution: Condition on the vector of sample means on control.

# Multi-arm platform trials



# Pre-planned adding

(Greenstreet et al., 2024)

- Timing of adding an arm is known
- Same control group and endpoint
- Allocation ratio pre and post adding fixed (but possibly different)

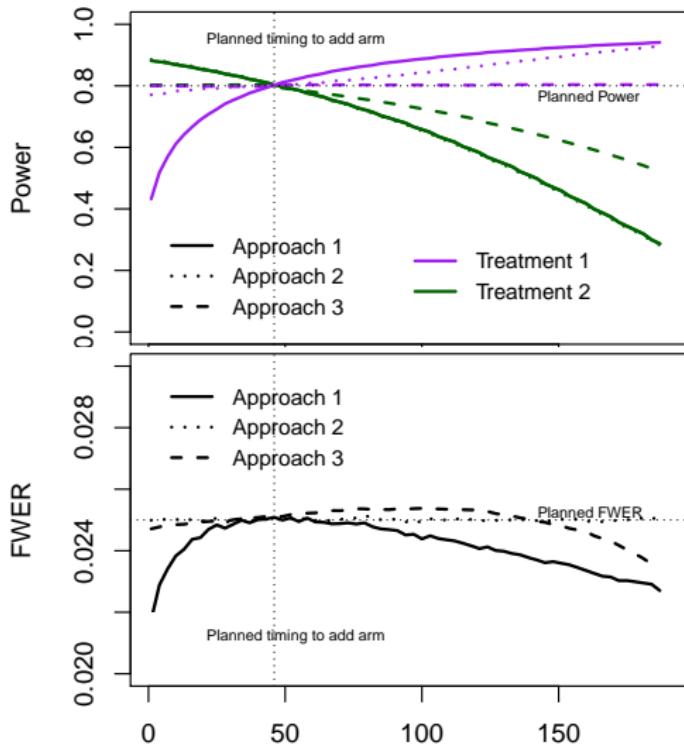
# Pre-planned adding

(Greenstreet et al., 2024)

- Timing of adding an arm is known
- Same control group and endpoint
- Allocation ratio pre and post adding fixed (but possibly different)

⇒ Same ideas as before can be used to find design

# Incorrect timing



A1: Move interim to when treatment is added but keep boundaries

A2: Move interim to when treatment is added and update boundaries

A3: Keep interim

When the second active treatment is added

# Pre-planned adaptive tests

## Advantages of pre-planned adaptive tests

1. Sufficient statistics
2. Sample size calculations
3. Confidence intervals always consistent with test decision

# Pre-planned adaptive tests

## Advantages of pre-planned adaptive tests

1. Sufficient statistics
2. Sample size calculations
3. Confidence intervals always consistent with test decision

**BUT** arm adding and treatment selection is typically more complex than a pre-planned rule.

## Conditional error (König et al, 2008)

- The conditional error,  $A(Y)$ , is the maximal probability of rejecting  $\mathcal{H}$  with the original test, conditional on the interim data  $Y$
- $B(Y)$  is the conditional error for a new test following an unplanned adaption.
- If  $B(Y) \leq A(Y)$  the new test controls the FWER.
- Can be used to account for ad-hoc design changes

Initial design:

- $J = 3, K = 2$
- $n = 10$  patients per arm per stage
- Triangular bounds

Modification:

- At first interim
- add 2 arms

# Operating characteristics

	$\theta$	$P(\mathbf{R})$	$E_{\theta}(N)$	$\mathbb{P}_{\theta}(R_1)$	$\mathbb{P}_{\theta}(R_2)$	$\mathbb{P}_{\theta}(R_3)$	$\mathbb{P}_{\theta}(R_4)$
2 trials	(0, 0, 0, 0)	0.08	87	0.03	0.03	0.03	0.03
	( $\delta$ , 0, 0, 0)	0.98	87	0.93	0.02	0.03	0.03
	( $\delta$ , 0, $\delta$ , 0)	0.98	86	0.93	0.02	0.82	0.04
	( $\delta$ , $\delta$ , $\delta$ , $\delta$ )	0.98	84	0.81	0.81	0.77	0.77
Start new	(0, 0, 0, 0)	0.05	62 (+30)	0.02	0.02	0.02	0.02
	( $\delta$ , 0, 0, 0)	0.76	63 (+30)	0.75	0.01	0.01	0.01
	( $\delta$ , 0, $\delta$ , 0)	0.89	62 (+30)	0.67	0.02	0.67	0.02
	( $\delta$ , $\delta$ , $\delta$ , $\delta$ )	0.95	63 (+30)	0.64	0.64	0.64	0.64
Cond Error	(0, 0, 0, 0)	0.05	78	0.03	0.02	0.01	0.01
	( $\delta$ , 0, 0, 0)	0.92	71	0.91	0.00	0.01	0.01
	( $\delta$ , 0, $\delta$ , 0)	0.95	71	0.86	0.01	0.64	0.03
	( $\delta$ , $\delta$ , $\delta$ , $\delta$ )	0.98	76	0.77	0.77	0.61	0.61

**Table:**  $R_i$  is the event that we reject  $H_{0i}$ ,  $\mathbf{R}$  is the event to reject at least one  $H_{0i}$ ,  $E_{\theta}(N)$  is the expected sample size.

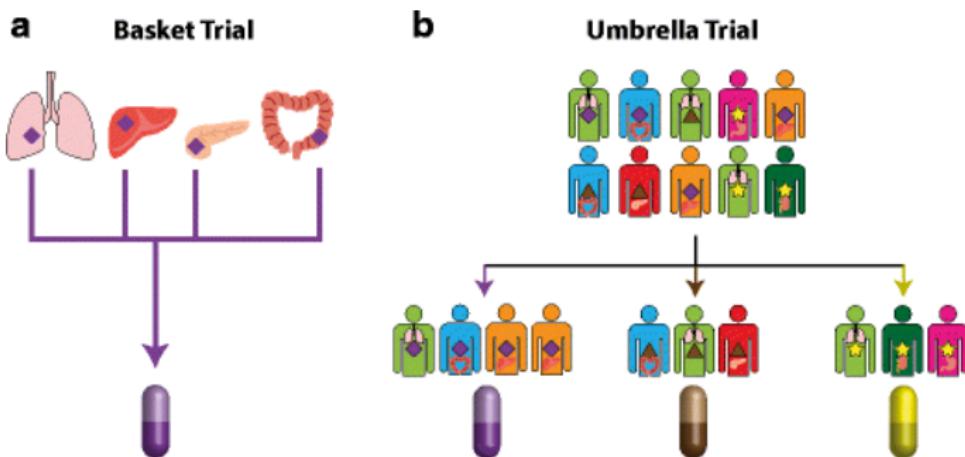
# Risk of wrong decision

- FWER is only one way to understand risk of wrong decisions
- Online FDR been explored in Robertson et al. (2023)
- Specific metrics have been suggested (e.g. Cui et al., 2023)

# Umbrella and Basket trials

Basket Trials Single treatment and single biomarker, different histologies placed in baskets

Umbrella Trials Single histology, multiple biomarkers each matched to treatments

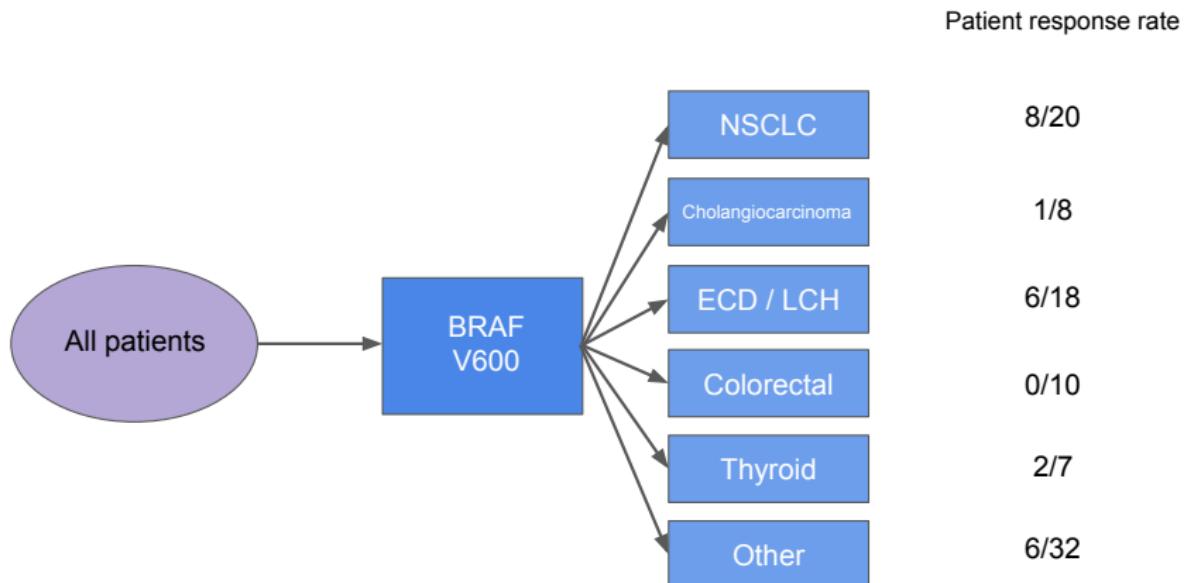


From Bui and Kummar (2018)

# Basket trials in oncology – An example

Hyman *et al.* (2015) reported a recent basket trial, which has been designed to evaluate the efficacy of vemurafenib in BRAF-V600.

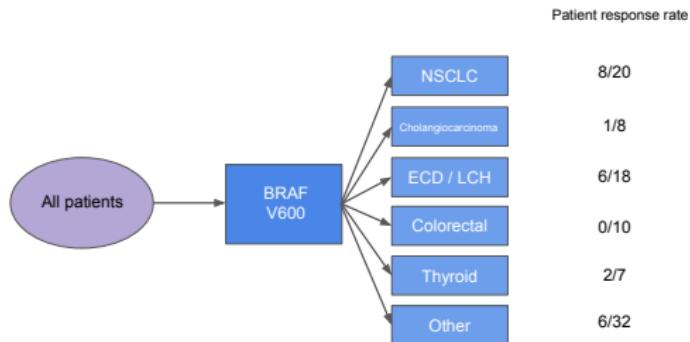
A total of 122 patients with BRAF-V600 mutations were enrolled, of which 95 entered the 6 modules.



# Borrowing of information between modules

With the **common genomic mutation** targeted by the investigational drug, one may expect:

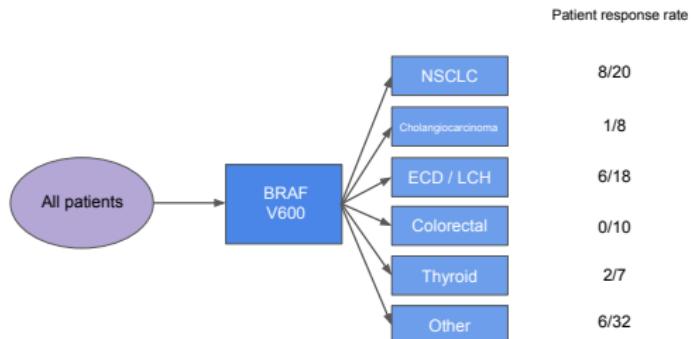
... some patient subgroups may **respond to the treatment similarly**.



# Borrowing of information between modules

With the **common genomic mutation** targeted by the investigational drug, one may expect:

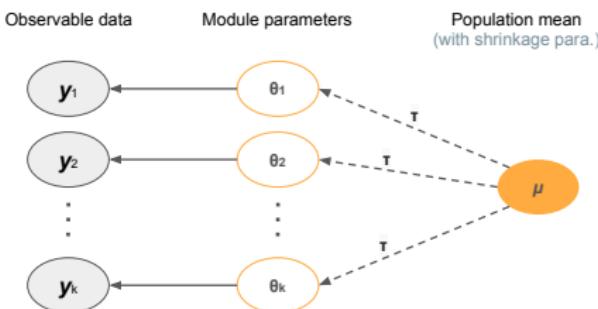
... some patient subgroups may **respond to the treatment similarly**.



## Potential analysis strategies:

- Stand-alone analyses
- Complete pooling
- Borrowing of information

# Bayesian hierarchical models

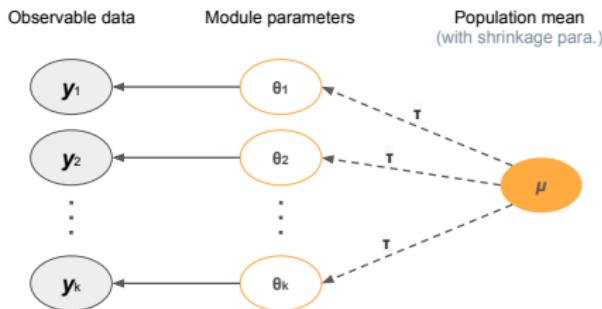


A hierarchical model for binomial data:

$$\begin{aligned}y_i | p_i, n_i &\sim \text{Binomial}(p_i, n_i) \\ \text{logit}(p_i) &= \theta_i \\ \theta_i | \mu, \tau &\sim N(\mu, \tau^2) \\ i &= 1, \dots, k.\end{aligned}$$

- ★ Hierarchical modelling assumes **exchangeability** (similarity) of the  $\theta_i$ s
- ★ The degree of borrowing is determined by  $\tau$ :
  - $\tau = 0 \rightarrow$  complete pooling of data from other modules;
  - $\tau = \infty \rightarrow$  no borrowing.
- ★ Very restrictive to suppose all  $\theta_i$ s will be shrunk towards one population mean

# Relaxing the exchangeability assumption – EXNEX



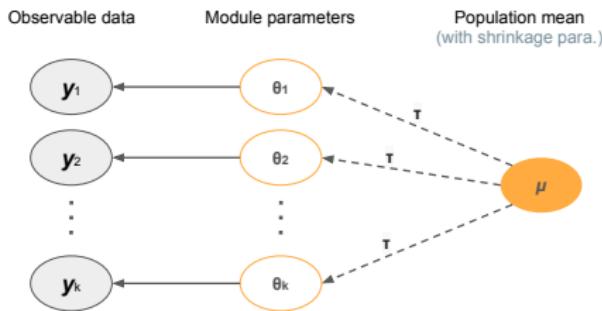
$$y_i | p_i, n_i \sim \text{Binomial}(p_i, n_i)$$

$$\text{logit}(p_i) = \theta_i$$

$$\theta_i | \mu, \tau \sim N(\mu, \tau^2)$$

$$i = 1, \dots, k.$$

# Relaxing the exchangeability assumption – EXNEX



$$y_i | p_i, n_i \sim \text{Binomial}(p_i, n_i)$$

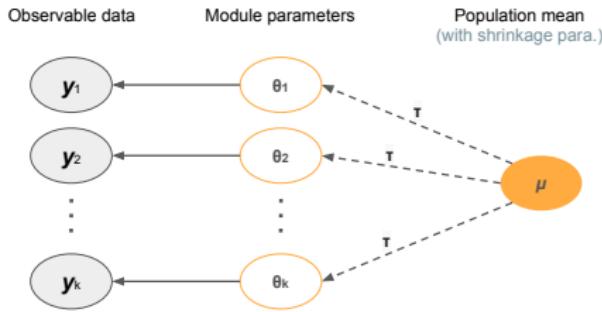
$$\text{logit}(p_i) = \theta_i$$

$$\theta_i | \mu, \tau \sim N(\mu, \tau^2)$$

$$i = 1, \dots, k.$$

... with probability of 1!

# Relaxing the exchangeability assumption – EXNEX



$$y_i | p_i, n_i \sim \text{Binomial}(p_i, n_i)$$

$$\text{logit}(p_i) = \theta_i$$

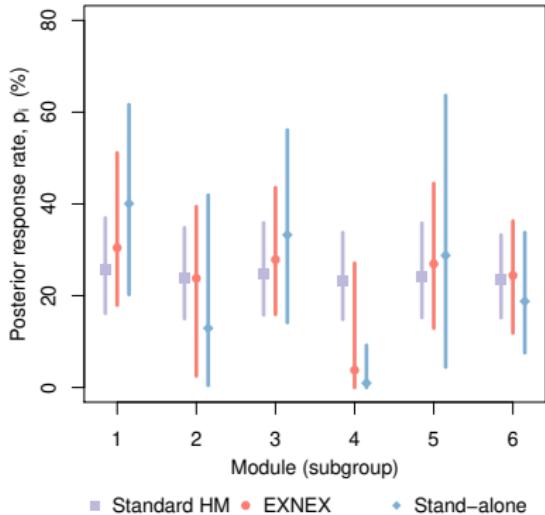
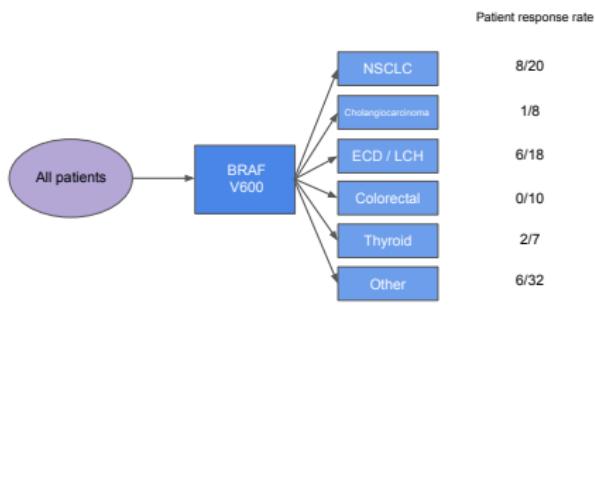
$$\theta_i | \mu, \tau \sim N(\mu, \tau^2)$$

$$i = 1, \dots, k.$$

... with probability of 1!

- ★ Neuenschwander *et al.* (2016) propose an extension to allow for **non-exchangeability**
  - ▶ **EX:**  $\theta_i | \mu, \tau \sim N(\mu, \tau^2)$  with probability of  $w_i$
  - ▶ **NEX:**  $\theta_i \sim N(m_i, s_i^2)$  with probability of  $1 - w_i$

# Comparing different analysis models



# Discussion

- In adaptive platform trials understanding risk of incorrect decisions is key
- Information borrowing a key feature of Basket and Umbrella trials
  - ▶ Should consistently be incorporated from design (Zheng et al, 2023) to analysis e.g. Zheng & Wason (2022)
- No one-size fits all solution for Master protocols, but case by case considerations necessary

# Literature (1)

- Bui NQ, Kummar S (2018) Evolution of early phase clinical trials in oncology. *Journal of Molecular Medicine*. **96**(1):31-38.
- Hyman D, Puzanov I, Subbiah V, et al. (2017) Vemurafenib in multiple nonmelanoma cancers with BRAF V600 mutations. *New England Journal of Medicine*. **373**(8):726-736.
- Magirr D, Jaki T, Whitehead J (2012) A generalized Dunnett test for multi-arm multi-stage clinical studies with treatment selection. *Biometrika*, **99**(2), 494–501.
- Neuenschwander B, Wandel S, Roychoudhury S, et al. (2016) Robust exchangeability designs for early phase clinical trials with multiple strata. *Pharmaceutical Statistics*. **15**(2):123-134.
- Bretz F, Koenig F. Commentary on Parker and Weir. (2020) *Clinical Trials*. **17**(5):567-569.
- Burnett T, Jaki T (2022) Making all pairwise comparisons in multi-arm multi-stage clinical trials without control treatment. Submitted.
- Burnett T, König F, Jaki T. (2024) Adding experimental treatment arms to Multi-Arm Multi-Stage platform trials in progress. *Statistics in Medicine*. **43**(18):3447-3462.
- Cui X, Liu Y, Schneider J, Tian H, Wang B, Hsu JC. (2023) Statistical Principles for Platform Trials. *arXiv:2302.12728*.
- Greenstreet P, Jaki T, Bedding A, Harbron C, Mozgunov P. (2024) A multi-arm multi-stage platform design that allows pre-planned addition of arms while still controlling the family-wise error. *Statistics in Medicine*. **43**(19):3613-3632.

## Literature (2)

- König F, Brannath W, Bretz F, Posch M (2008) Adaptive Dunnett Tests for Treatment Selection. *Statistics in Medicine* **27**, 1612-1625.
- Magirr D, Stallard N, Jaki T (2014) Flexible sequential designs for multi-arm clinical trials. *Statistics in Medicine*. **33**, 3269-3279.
- Parker RA, Weir CJ. Non-adjustment for multiple testing in multi-arm trials of distinct treatments: rationale and justification. *Clinical Trials*. **17**(5):562-566.
- Robertson DS, Wason J, König F, Posch M, Jaki T. (2023) Online error control for platform trials. *Statistics in Medicine*. **42**(14):2475-2495.
- Zheng H, Wason JM. (2022) Borrowing of information across patient subgroups in a basket trial based on distributional discrepancy. *Biostatistics*. **23**(1):120-135.
- Zheng H, Jaki T, Wason JM. (2023) Bayesian sample size determination using commensurate priors to leverage preexperimental data. *Biometrics*. **79**(2):669-683.