2022/12/2 AMED Platform Trial Symposium

Statistical Issues in Platform Trial

Department of Biostatistics and Informatics, The University of Tokyo

Kohei Uemura

Statistical issues in platform trial

- 1. Multiplicity problem
- 2. Response adaptive randomization (RAR)
- 3. Utilization of non-concurrent (NC) control

How should we consider multiplicity problem ?

- If we regard a platform trial as a single clinical trial
  - Need to control the **platform-wise error rate** as usual trial-wise ?
- If we consider a platform in which
  - multi- hypotheses(arms) to be accessed in independent trials
  - are accessed in a platform
  - Control of FWER may **NOT** be needed…

JOURNAL OF BIOPHARMACEUTICAL STATISTICS
2020, VOL. 30, NO. 6, 1077–1090
https://doi.org/10.1080/10543406.2020.1821703

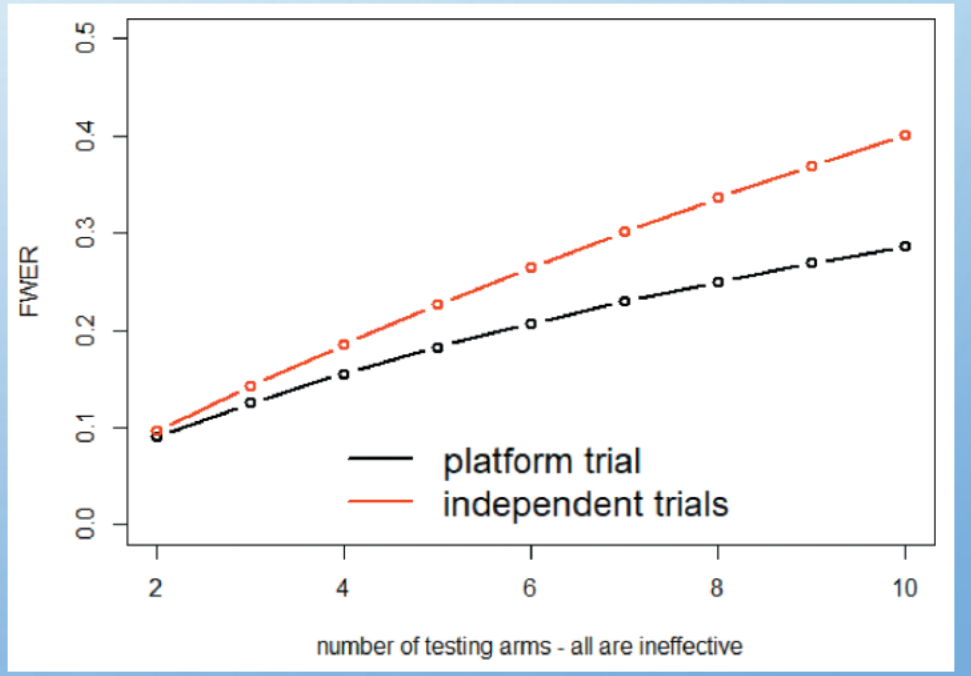
Taylor & Francis
Taylor & Francis Group

Check for updates

Multiplicity issues for platform trials with a shared control arm

Xiaofei Bai, Qiqi Deng, and Dacheng Liu

Department of Biostatistics and Data Sciences, Boehringer-Ingelheim Pharmaceutical Inc., Ridgefield, Connecticut, USA



How should we consider multiplicity problem ?

- FWER of platform trial is **lower** than independent trials
  - **Due to correlation** among test statistics via shared control arm

JOURNAL OF BIOPHARMACEUTICAL STATISTICS
2020, VOL. 30, NO. 6, 1077–1090
https://doi.org/10.1080/10543406.2020.1821703

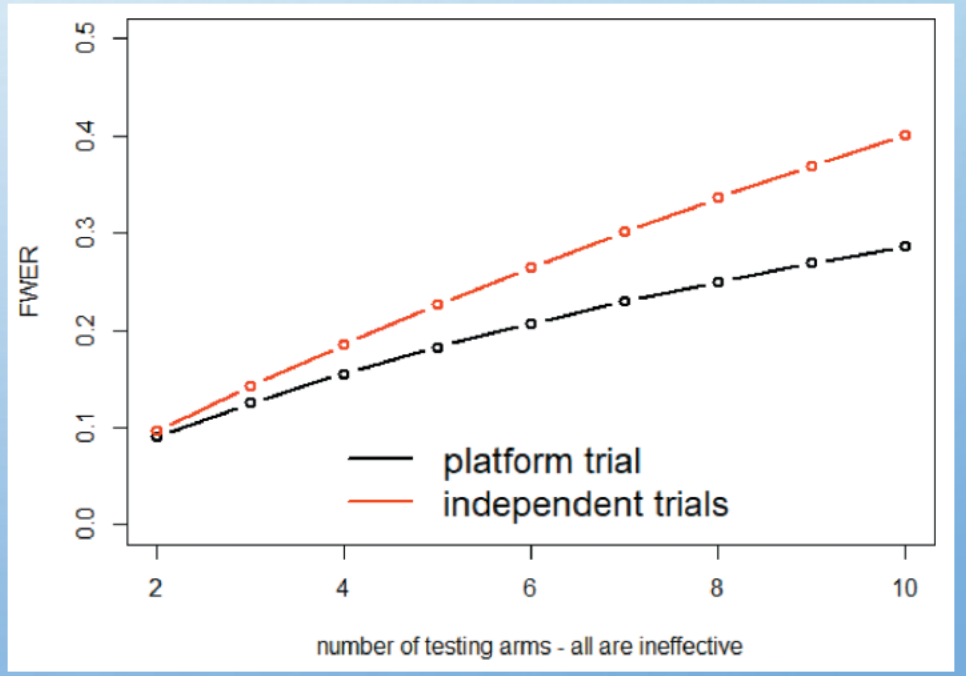
Taylor & Francis
Taylor & Francis Group

Check for updates

Multiplicity issues for platform trials with a shared control arm

Xiaofei Bai, Qiqi Deng, and Dacheng Liu

Department of Biostatistics and Data Sciences, Boehringer-Ingelheim Pharmaceutical Inc., Ridgefield, Connecticut, USA



○○-wise error rate

- If it is a situation of multi- hypotheses(arms, **aims**) to be **independently** accessed
  - Discussion and conclusions about results should be conducted separately
  - If a conclusion for analysis result is like that arm A and C were effective among all tried arms, I feel need of FWER control

○○-wise error rate

- It depends on what is the main constituent of platform
  - State-sponsored infrastructure, Research Organization Groups, a global company, …
- ○○-wise error rate
  - ○○ may not be Platform as the very infra. for long-lasting
  - ○○ may be **domain of treatment**, **state of disease**, **current wave of epidemic**, …

Is RAR useful?

- Korn & Freidlin (2011)
- Overall P(responders)%
  - ◆33.2 or 33.7% (RAR) vs 30% (1:1)
- More # of Non-responders
  - ◆93.5, 92.9 vs 92.4
- More Ave. Sample Size
  - ◆83.7 vs 78.4(Table 3 with Efficacy stop.)

VOLUME 29 · NUMBER 6 · FEBRUARY 20 2011
JOURNAL OF CLINICAL ONCOLOGY | STATISTICS IN ONCOLOGY

Outcome-Adaptive Randomization: Is It Useful?

Edward L. Korn and Boris Freidlin

Table 2. Average Proportion of Responders, No. of Nonresponders, and Overall Proportion Treated on the Experimental Arm for Various Randomized Phase II Trial Designs, Some of Which Use Adaptive Randomization

Response Rates by Arm		Fixed Sample Size				Adaptive Randomization (N = 140)					
		1:1 (n = 132)		2:1 (n = 153)		Capped at 80% Assignment Probability			Capped at 90% Assignment Probability		
Control Arm	Experimental Arm	P (responders) %	No. of Nonresponders	P (responders) %	No. of Nonresponders	P (responders) %	No. of Nonresponders	Overall % Treated on Experimental Arm	P (responders) %	No. of Nonresponders	Overall % Treated on Experimental Arm
0.2	0.2	20.0	105.6	20.0	122.4	20.0	112.0	50.0	20.0	112.0	50.0
0.2	0.3	25.0	99.0	26.6	112.2	26.0	103.6	59.7	26.0	103.6	60.3
0.2	**0.4**	**30.0**	**92.4**	**33.3**	**102.0**	**33.2**	**93.5**	**66.2**	**33.7**	**92.9**	**68.2**
0.2	0.5	35.0	85.8	40.0	91.8	41.0	82.6	69.9	42.1	81.1	73.6

NOTE. Adaptive randomization uses the method of Thall and Wathen[12] but with no early stopping. One-sided type 1 error = 10%, power = 90% at 20% *v* 40% response rates; results based on 500,000 simulations. Characteristics of trial designs corresponding to the trial alternative hypothesis are in bold type. P (responders) % is the average proportions of responders given as a percentage.

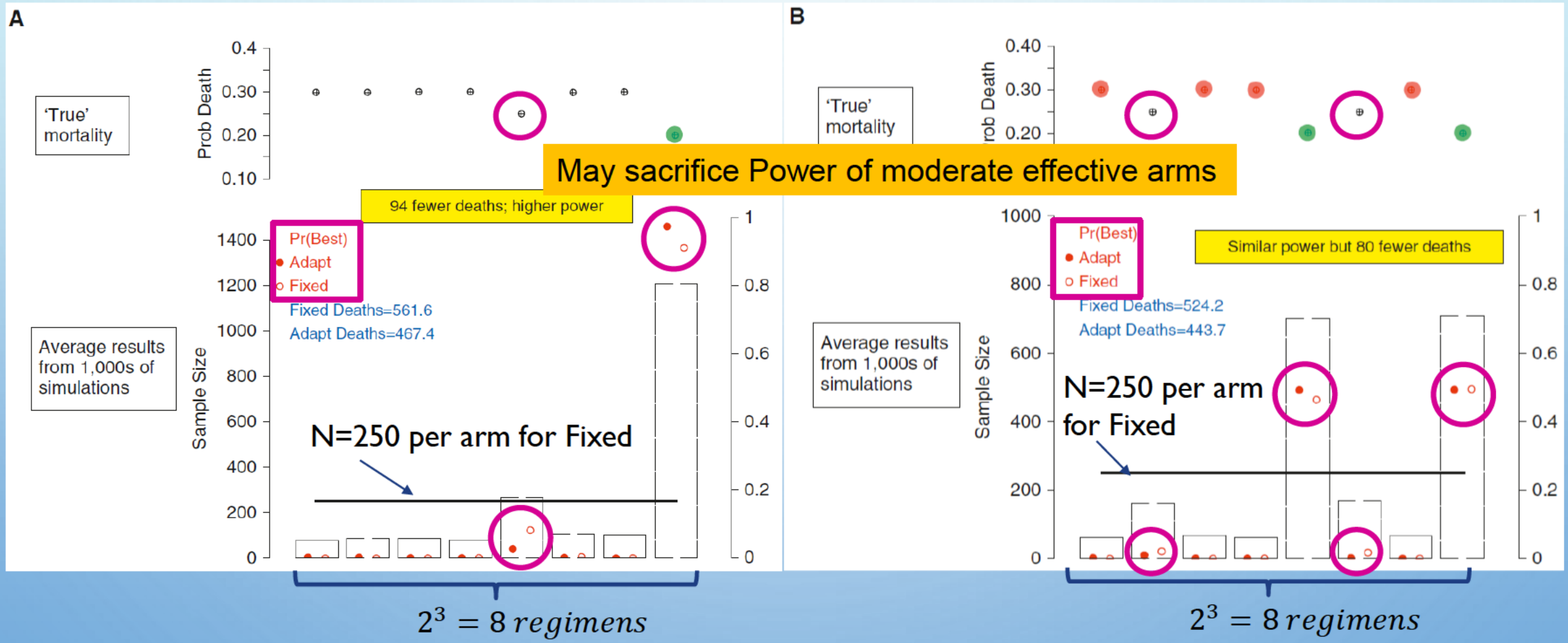
Table 3. Average Sample Size, Proportion of Responders, and No. of Nonresponders for Fixed 1:1 and Adaptive Randomized Phase II Trial Design

Response Rates by Treatment Arm		Fixed 1:1 (maximum sample size = 190)			Adaptive Randomization Capped at 80% Assignment Probability (maximum sample size = 208)		
Control Arm	Experimental Arm	Average sample size	P (responders) %	No. of Nonresponders	Average sample size	P (responders) %	No. of Nonresponders
0.2	0.2	177.9	20.3	142.3	194.3	20.2	155.5
0.2	0.3	135.2	25.9	101.4	147.6	26.3	109.7
0.2	**0.4**	**78.4**	**31.4**	**54.8**	**83.7**	**32.1**	**57.3**
0.2	0.5	43.3	36.6	28.2	45.3	37.1	28.8

NOTE. Adaptive randomization uses the method of Thall and Wathen[12]. Trials are stopped early for superiority of the experimental treatment if $P(E > C) > 0.984$. One-sided type 1 error = 10%; power = 90% at 20% *v* 40% response rates; results based on 500,000 simulations. Characteristics of trial designs corresponding to the trial alternative hypothesis are in bold type. P (responders) % is the average proportions of responders given as a percentage.

Trial Simulation : REMAP-CAP Design vs. Standard Design
3 domains with each 2 interventions

Ann Am Thorac Soc Vol 17, No 7, pp 879–891, Jul 2020 (Received in original form March 3, 2020; accepted in final form April 8, 2020)



Response Adaptive Randomization

- **Early trial period** with little information and **frequent update of allocation ratio**
  - May make **prob. of best arm unstable** and can not allocate to right arm for individual patient not for average population (Proschan & Evans, CID2020)


Clinical Infectious Diseases

INVITED ARTICLE

IDSA Infectious Diseases Society of America hivma hiv medicine association

INNOVATIONS IN DESIGN, EDUCATION AND ANALYSIS (IDEA): Victor De Gruttola and Scott R. Evans, Section Editors

Resist the Temptation of Response-Adaptive Randomization

Michael Proschan[1,©] and Scott Evans[2]

[1]Mathematical Statistician, Biostatistics Research Branch, National Institute of Allergy and Infectious Diseases, Rockville, Maryland, USA, and [2]Department of Biostatistics and Bioinformatics; Director, Biostatistics Center, Milken Institute School of Public Health, George Washington University, Washington, DC, USA

  - **Burn-in period** and **Stage-wise update** of allocation ratio may work better(One can conduct a stratified analysis to adjust time trend bias)

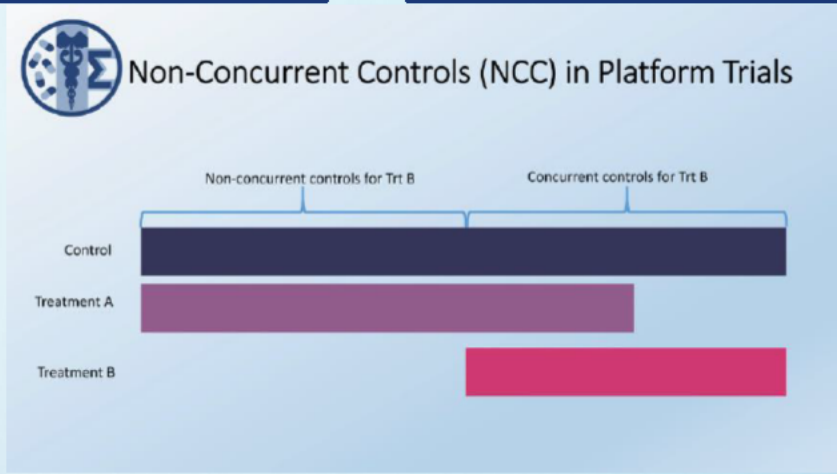
Response Adaptive Randomization

- Performance of RAR strongly **depends on true scenario**
  - **# of effective arms** and **# of in-effective arms**
  - **Small difference** between arms make RAR **not to work well**
  - Power of **moderate effective** arms may be sacrificed

Response Adaptive Randomization

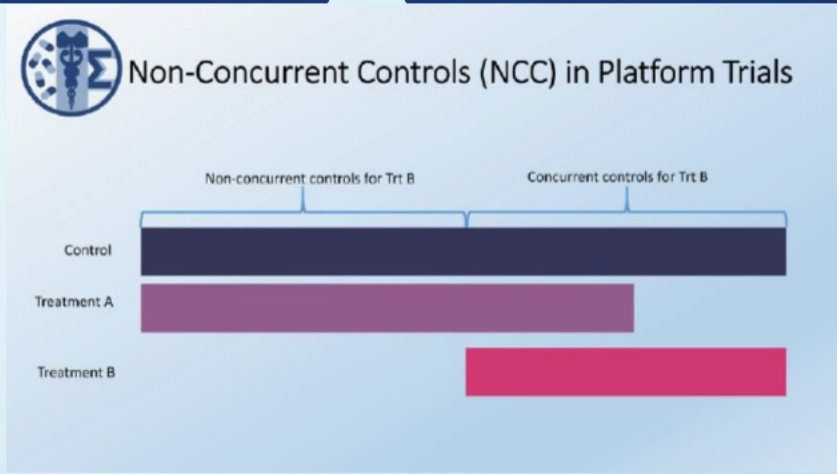
- Do not fit to **multi-aims** platform concept,
  - rather fit to **decide best** regimen or **screening** of promising arm in multiple candidates which probably include many ineffective arms
- **Ethical design option** to start RAR at interim stage
  - if considerable unbalance among investigational treatment arms (other than control arm) were observed

Open vs. Closed Platform



- Open platform trial
  - **Allows to add** on an arm at an interim
  - Need adaptation of allocation (not equal to RAR)
  - Fit to situations where it takes **a certain amount of time to recruit** and additional therapeutic candidates are necessarily added **along the long way**

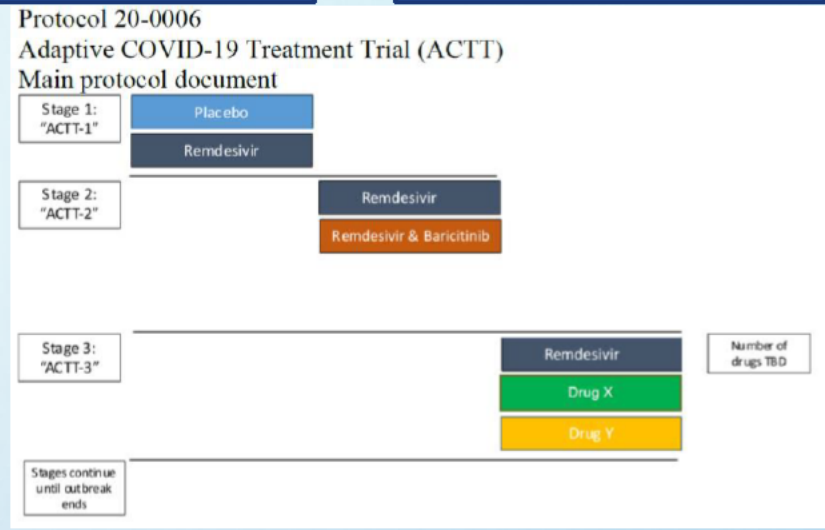
Open vs. Closed Platform



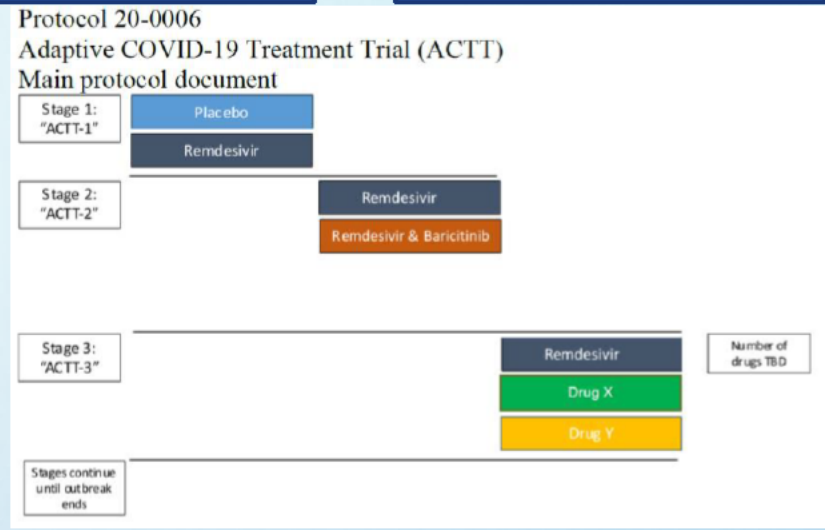
- Open platform trial
  - **Non-concurrent(NC)** control can **NOT basically** be compared to an added arm in a regulatory setting
    - ✓which leads to **less power** otherwise randomization continue after existing arms reached to the sufficient size
  - On the other hand, more efficient than to wait and **start a new trial** for an added arm

Open vs. Closed Platform

- Closed platform trial
  - **NOT allow** to add on an arm at an interim
  - **No occurrence** of non-concurrent control
  - Can **efficiently** compare to control
  - Easy to change control arm

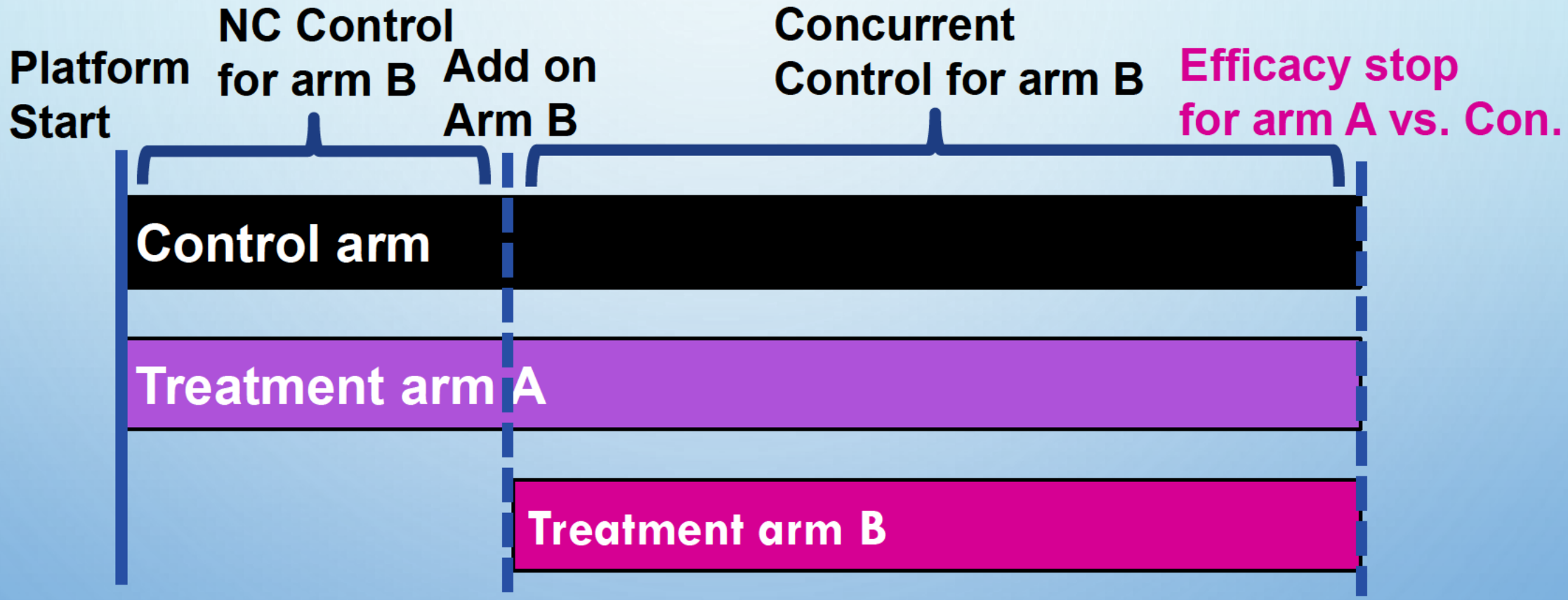


Open vs. Closed Platform

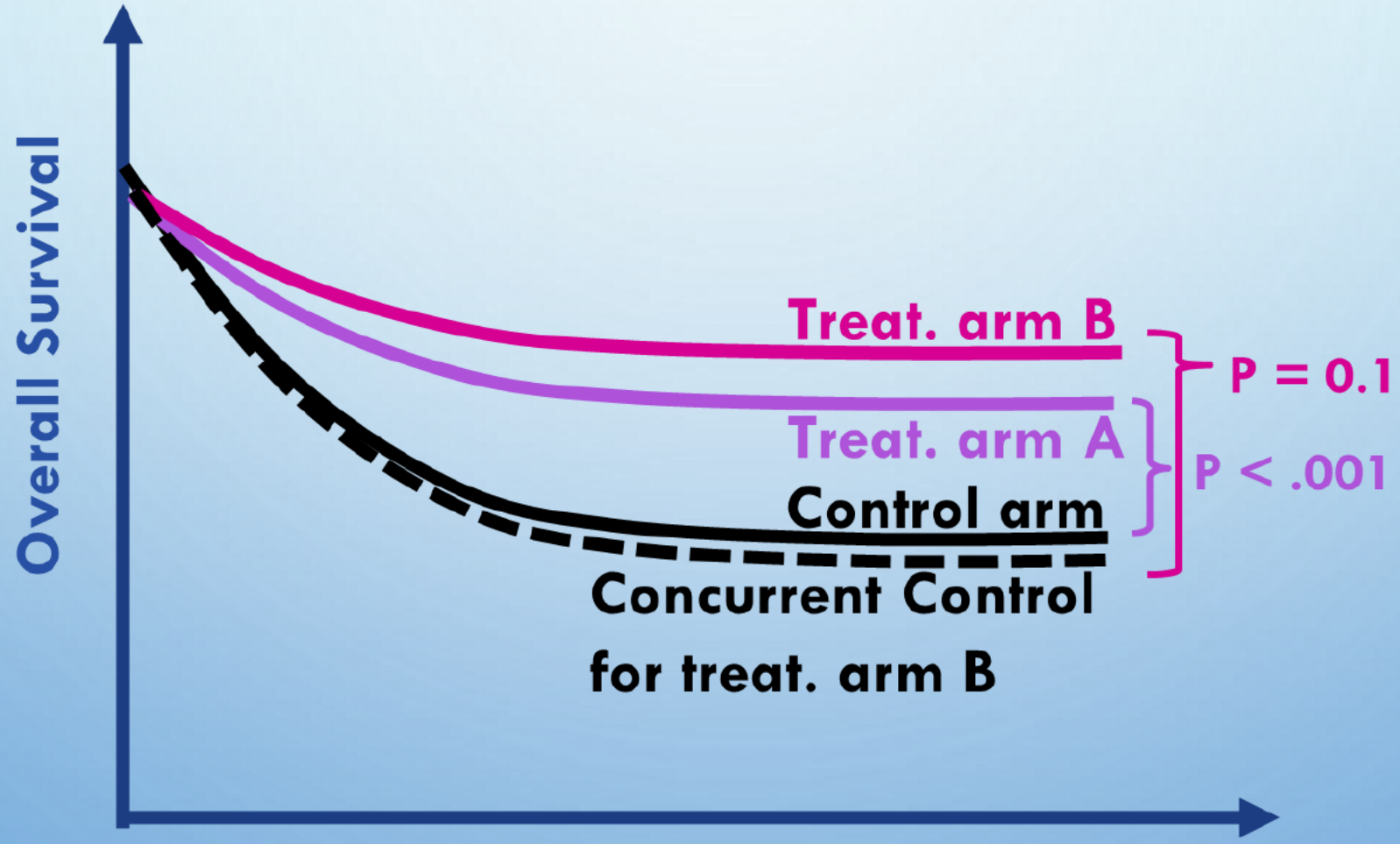


- Closed platform fit to situations where
  - Candidates have been **narrowed down** at the start
  - It takes only a **short time** to recruit
  - One **can start another stage** when additional therapeutic candidates are added along the way

What can we do **if concurrent control stopped** as arm A showed superiority ?(An issue of open platform)

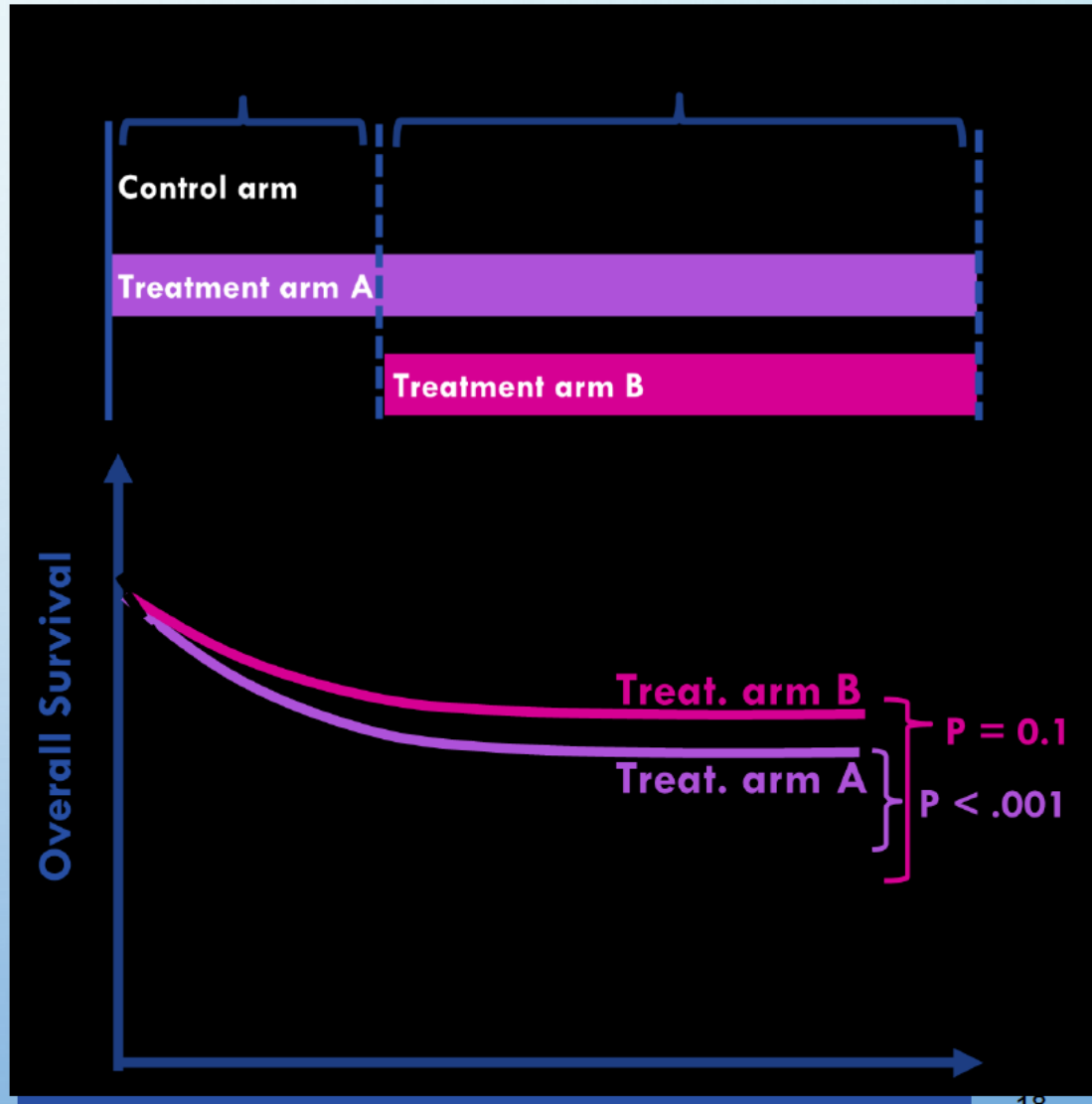


Treatment B indicate **promising trend** but do not show a definitive conclusion



What can we do if concurrent control stopped ?

- **Extend** the follow up
  - Time-to-event endpoint as OS
- **Extend** the recruit of arm A&B
  - If OS curve reach to plateau
  - And we can not extend Cont. arm
  - **Switch to Non-inferiority** of B vs A
- **Utilize NC** Control data
  - If Control & NC are similar in trend

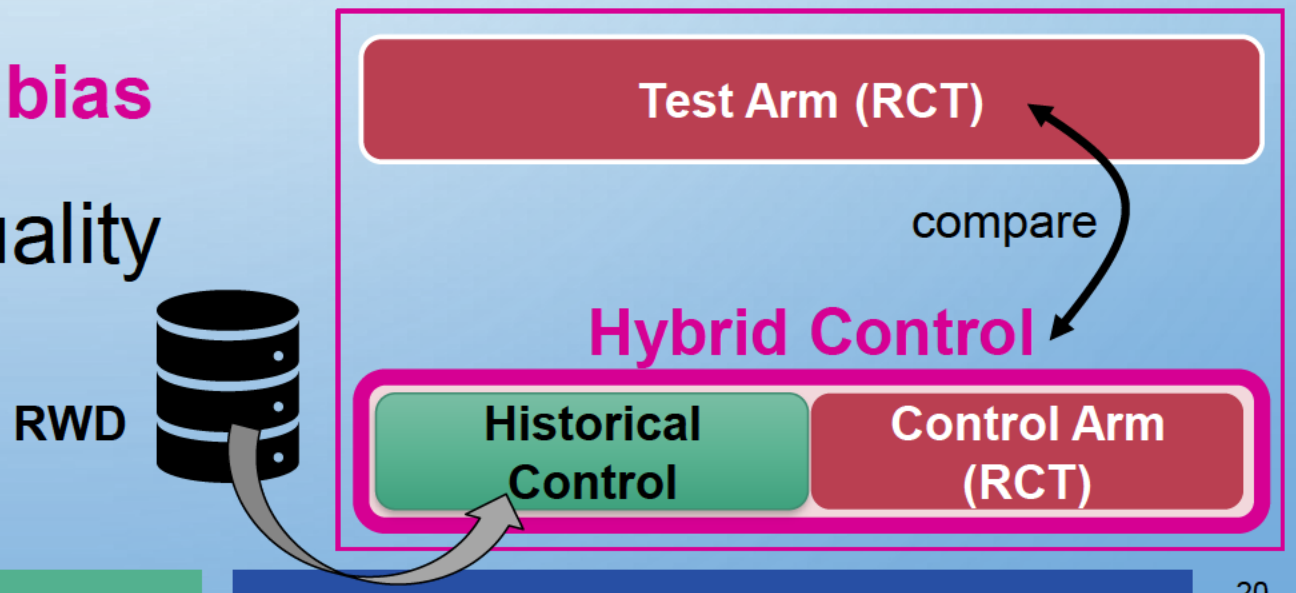


How to utilize NC control data ?

- The lack of head-to-head randomization may introduce **measured and unmeasured confounding** effects
  - Prognostic factors may not be balanced
  - Patients and investigators may not be blinded
  - Unmeasured confounding effects may be **related to time trends**
- Hybrid Control Approach
- Model based approach

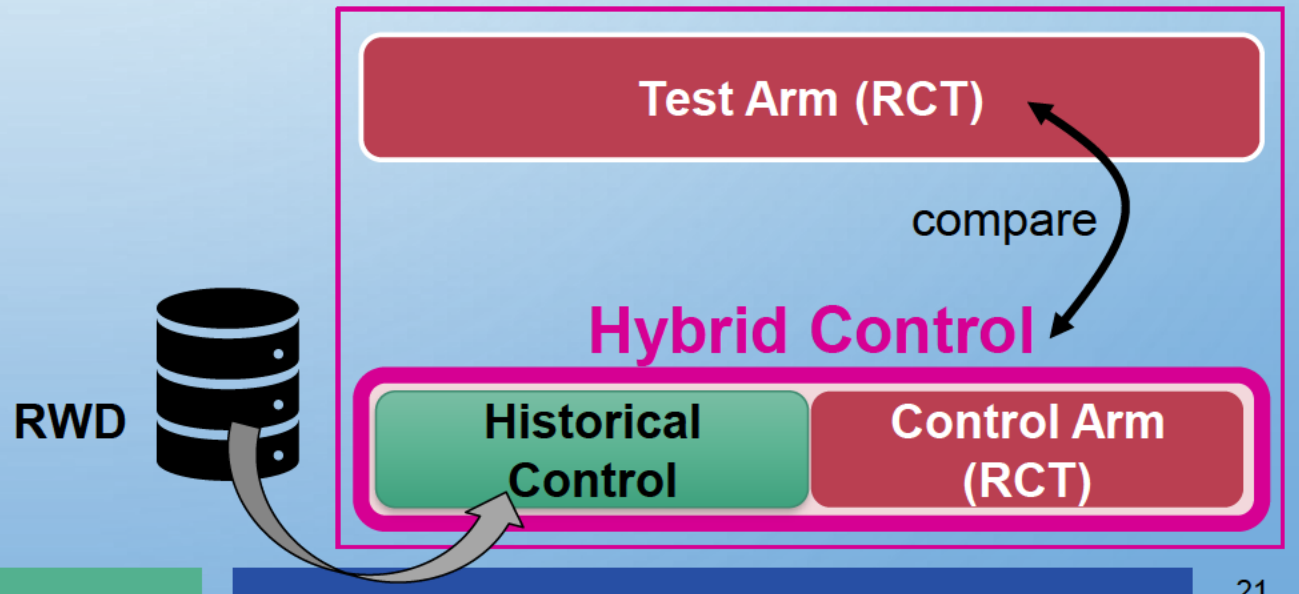
Hybrid Control Approach

- Can adapt to **lack of sample size** of randomized control
  - Utilizing some historical data
- NC control can be **a good historical control**
  - If **time trend** is **minimized** and there is **no operational bias**
- Patient population, data quality
  - **can be similar** to randomized arms



Hybrid Control Approach

- Collected data **within the same framework**
  - recruiting centers
  - inclusion/exclusion criteria
  - endpoint assessments, etc.



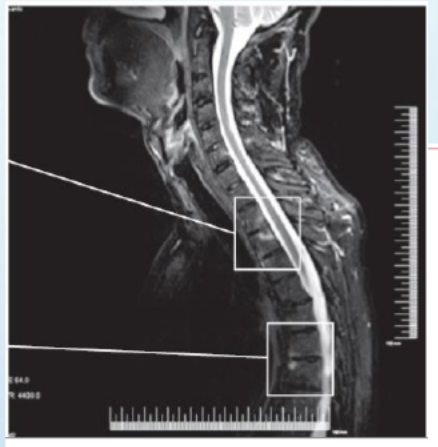
Application of Hybrid Control@FDA

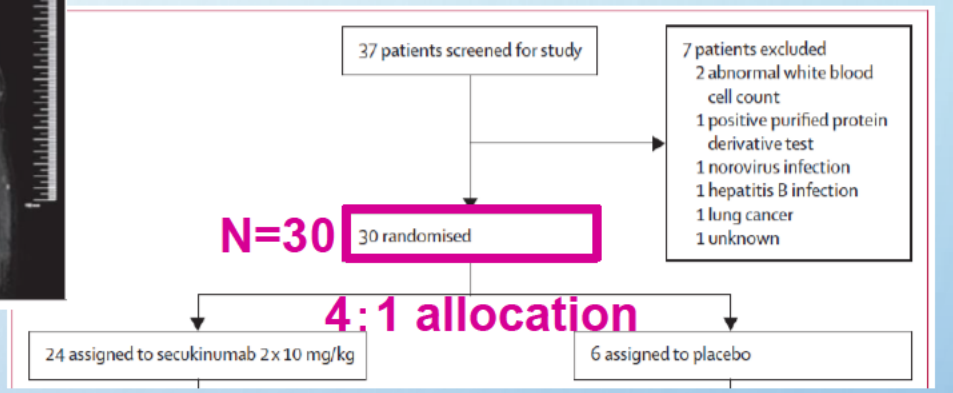
Lancet 2013; 382: 1705–13

Anti-interleukin-17A monoclonal antibody secukinumab in treatment of ankylosing spondylitis: a randomised, double-blind, placebo-controlled trial

Dominique Baeten, Xenofon Baraliakos, Jürgen Braun, Joachim Sieper, Paul Emery, Désirée van der Heijde, Iain McInnes, Jacob M van Laar, Robert Landewé, Paul Wordsworth, Jürgen Wollenhaupt, Herbert Kellner, Jacqueline Paramarta, Jiawei Wei, Arndt Brachat, Stephan Bek, Didier Laurent, Yali Li, Ying A Wang, Arthur P Bertolino, Sandra Gsteiger, Andrew M Wright, Wolfgang Hueber

- A RCT for Anti-IL17 monoclonal antibody (secukinumab) in treat. of **ankylosing spondylitis**
- **ASAS20% response rate (PE)**
  - **Anti-IL17：14/23(60.9%)**
  - **placebo：1/6(16.7%)**
    ⇒Hybrid placebo Control：24.5%
  - Borrowing historical placebo**11/43例(25.6%)**
    - ✓Meta-analysis of **8 RCT(n=533)**
  - Diff vs P: **34.7%[11.5-56.4%]**





	Responders, n (%)	Response rate*	Difference vs placebo†	95% credibility interval†	Probability
Secukinumab‡§	14 (60·9%)	59·2%	34·7%	11·5–56·4%	99·8%
Placebo	1 (16·7%)	24·5%	..	..	..

ASAS=Assessment of SpondyloArthritis international Society criteria. *Means from the posterior beta (0·5+x, 1+n−x) distribution for secukinumab and beta (11+x, 32+n−x) distribution for placebo, where x represents the number of responders and n−x represents the number of non-responders in the corresponding treatment group. †Difference in response rates simulated from the posterior probability distributions of secukinumab and placebo. ‡Secukinumab: 2×10 mg/kg. §The efficacy dataset included only 23 of 24 patients in the secukinumab group, since one patient was excluded due to a dosing error.

Table 2: Primary endpoint Bayesian analysis of ASAS20 responders at week 6

Adjusting time trend bias via statistical modeling

- Bayesian **"time machine"** (Saville and Berry, 2016)
  - Platform trials such as I-Spy 2, GBM-AGILE, and Precision Promise, REMAP-CAP
  - **A parameter is smoothed** by Normal Dynamic Linear Model
- The NDLM allows for **borrowing** among effects of **adjacent time periods**
  - Pulling their estimates towards each other, and can **robustly handle** different trends over time

Bayesian hierarchical drift model using NDLM

- Time-indicating variable t is from 1 to T
  - **Response rate P_{jkT}** of subgroup j, treatment k at $t = T$ is modeled by a Bayesian hierarchical model
- For all $t < T$
  - $Y_{jkt} \sim Binomial(n_{jkt}, P_{jkt})$
- **Time effect** parameters θ_t are modeled with NDLM
  - $logit(P_{jkt}) = logit(P_{jkT}) + \theta_t$
  - $\theta_{t-1} \sim N(\theta_t, \tau), \tau \sim InvGamma(0.25, 0.1)$
  - Borrowing is controlled by the **drift parameter τ**

Benefits and drawbacks of using NC control data

Lee and Wason *BMC Medical Research Methodology* (2020) 20:165
https://doi.org/10.1186/s12874-020-01043-6

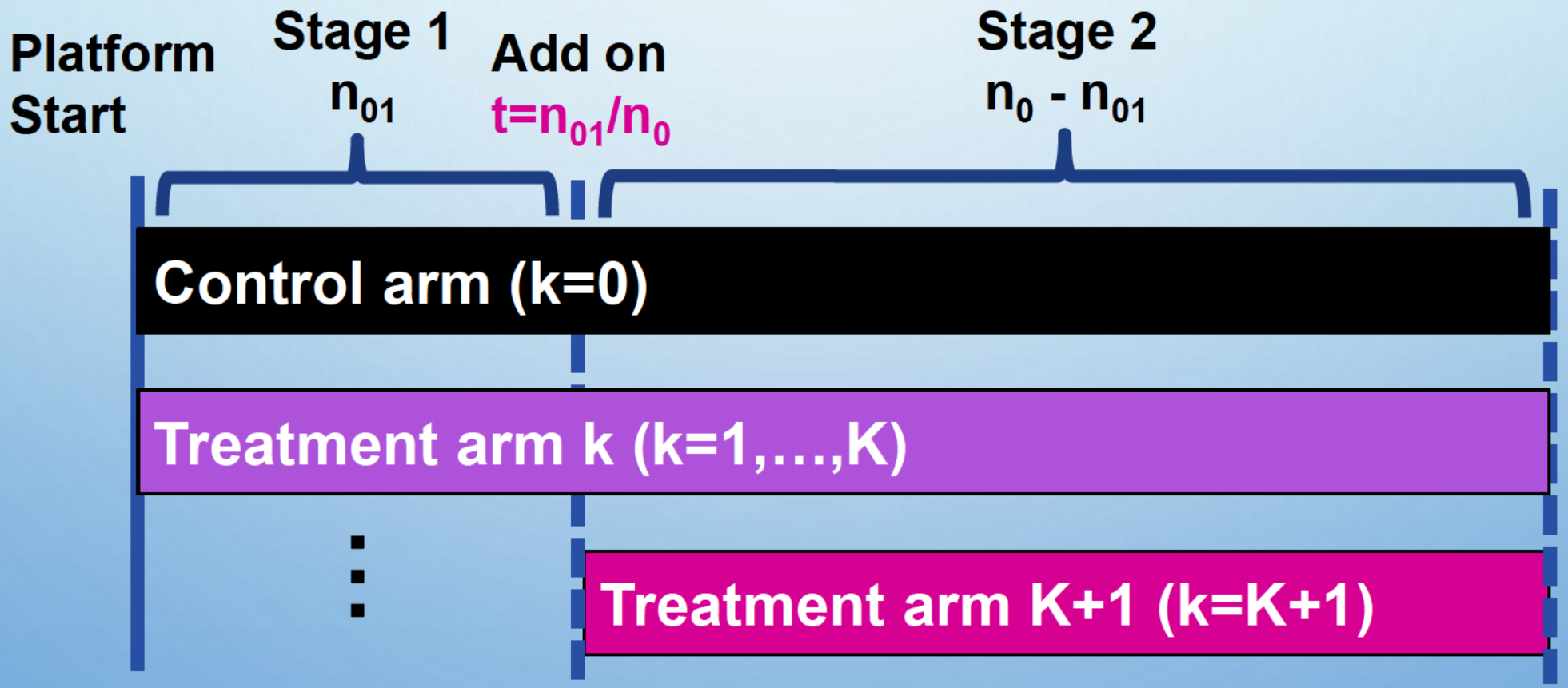
BMC Medical Research Methodology

RESEARCH ARTICLE Open Access

Including non-concurrent control patients in the analysis of platform trials: is it worth it?

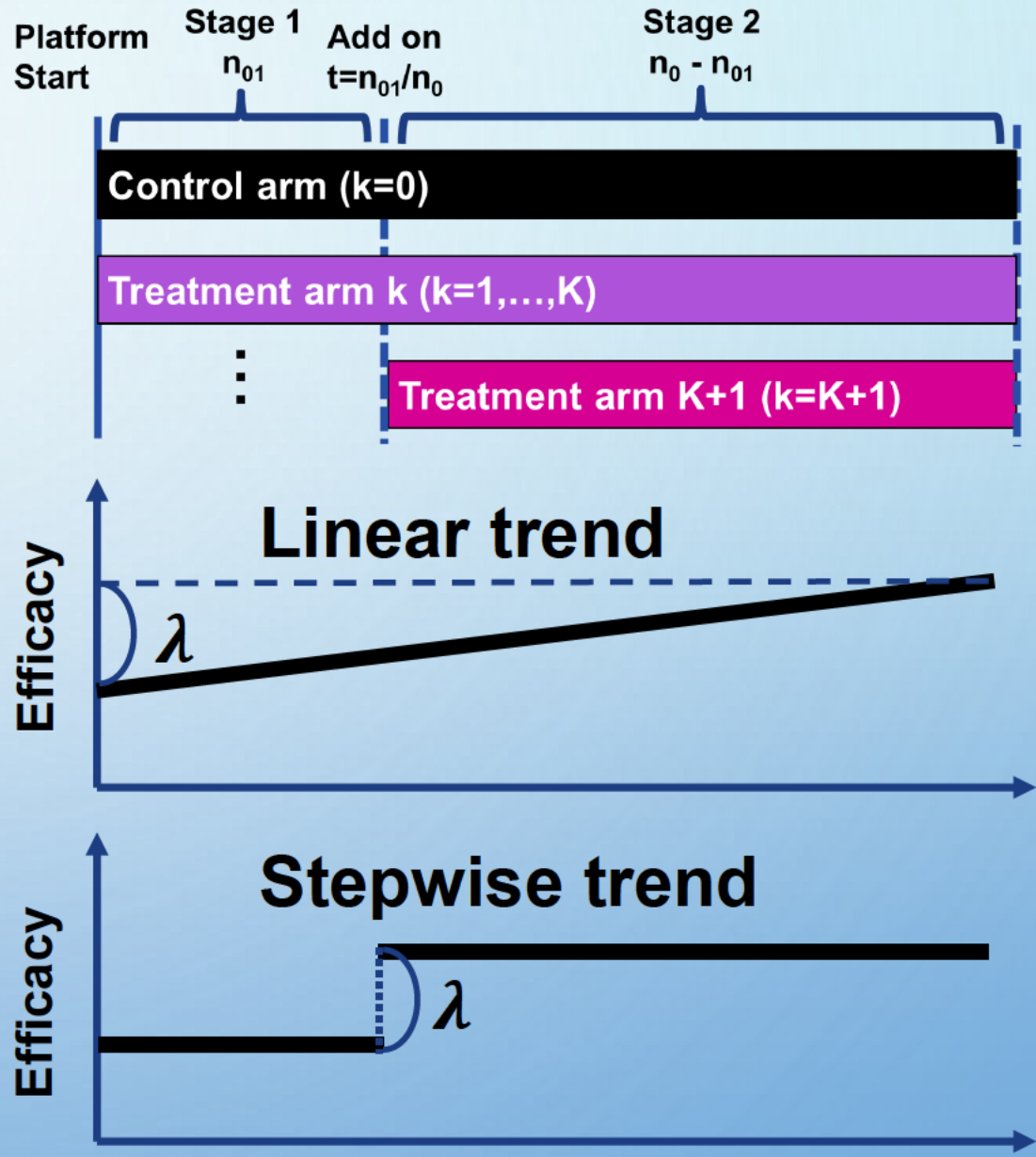
Kim May Lee[1*] and James Wason[1,2]

- Simple **two-stage** setting: outcome variable is **continuous**



Adjusting time trend bias

- Regression Model-based approach
  - Under two types of trend
- Utilize **all** treat. arms (k=0,…,K+1)
  - j: $subject\ number$
  - $M_{a1}: X_{jk} = \beta_k + \boldsymbol{\gamma \cdot j} + \epsilon_j,$
  - $M_{a2}: X_{jk} = \beta_k + \boldsymbol{\nu \cdot I(j \in stage\ 2)} + \epsilon_j$
- Use **only 2** arms (k=0, K+1)
  - $M_{b1}: X_{jk} = \beta_k + \boldsymbol{\gamma \cdot j} + \epsilon_j$
  - $M_{b2}: X_{jk} = \beta_k + \boldsymbol{\nu \cdot I(j \in stage\ 2)} + \epsilon_j$



Simulation Results: k=0, 1, 2 (2 arms + 1 arm)

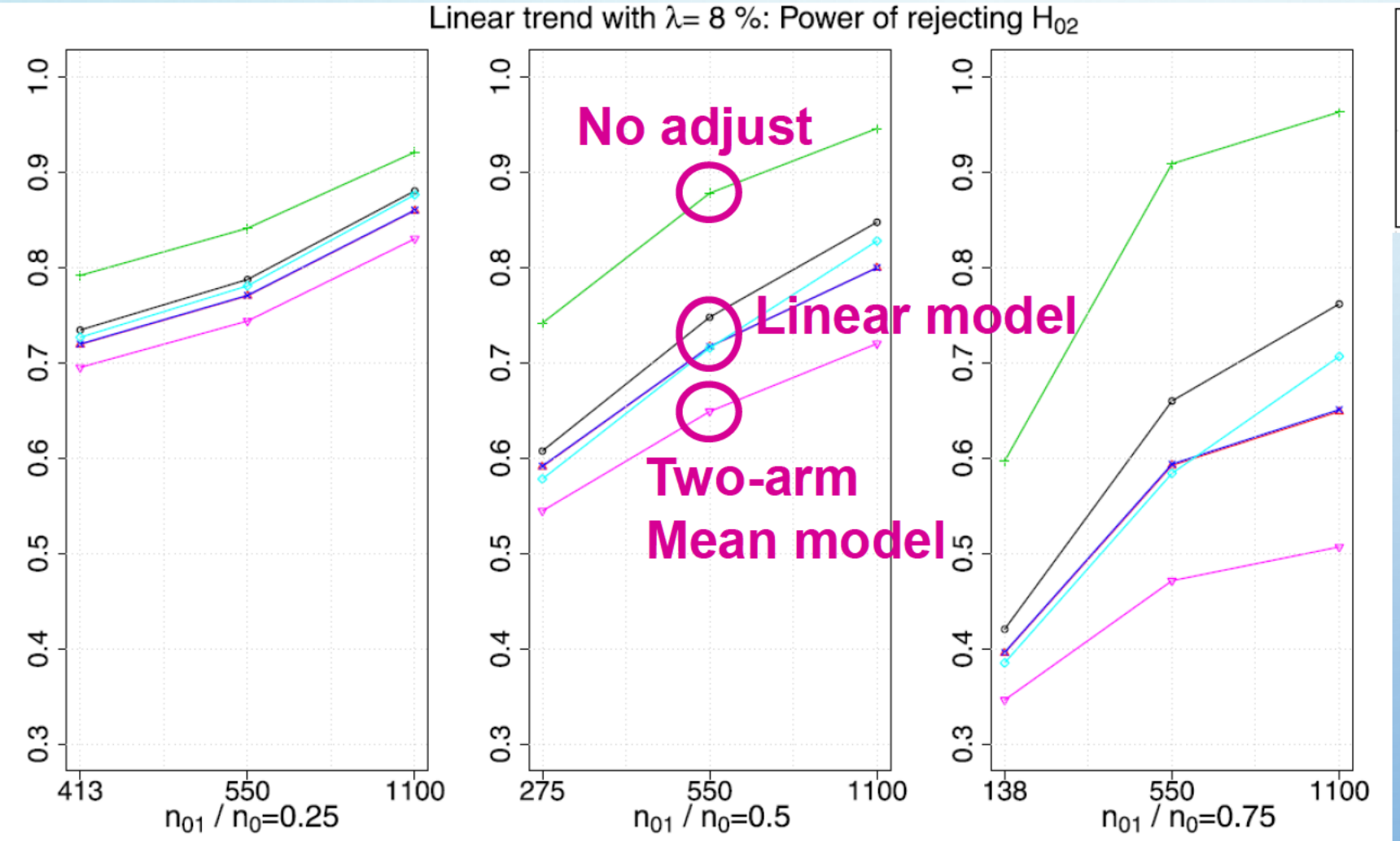
$$\delta = 0.15, \sigma^2 = 1, n_{01}/n_0 = \{0.25, 0.5, 0.75\}, n_{22} = \{n_{02}, n_0, 2n_0\}$$

Table 1 The maximum (median) absolute bias of the estimated difference in mean responses of the newly added treatment and the control arm when there is a trend. Values with -4 order of magnitude are set to zero

λ	WLS_{all}	WLS_{s2}	M_{a1}	M_{a2}	M_{b1}	M_{b2}
			Linear trend			
2%	0.007 (0.005)	0 (0)	0.001 (0)	0 (0)	0 (0)	0 (0)
4%	0.015 (0.010)	0 (0)	0.001 (0)	0 (0)	0 (0)	0 (0)
6%	0.022 (0.015)	0 (0)	0.001 (0)	0 (0)	0 (0)	0 (0)
8%	0.030 (0.020)	0 (0)	0.001 (0)	0 (0)	0 (0)	0 (0)
			Step trend			
2%	0.015 (0.010)	0 (0)	0.007 (0.004)	0 (0)	0.010 (0.005)	0 (0)
4%	0.030 (0.020)	0 (0)	0.015 (0.008)	0 (0)	0.019 (0.010)	0 (0)
6%	0.045 (0.030)	0 (0)	0.023 (0.012)	0 (0)	0.029 (0.015)	0 (0)
8%	0.060 (0.040)	0 (0)	0.031 (0.016)	0 (0)	0.038 (0.019)	0 (0)

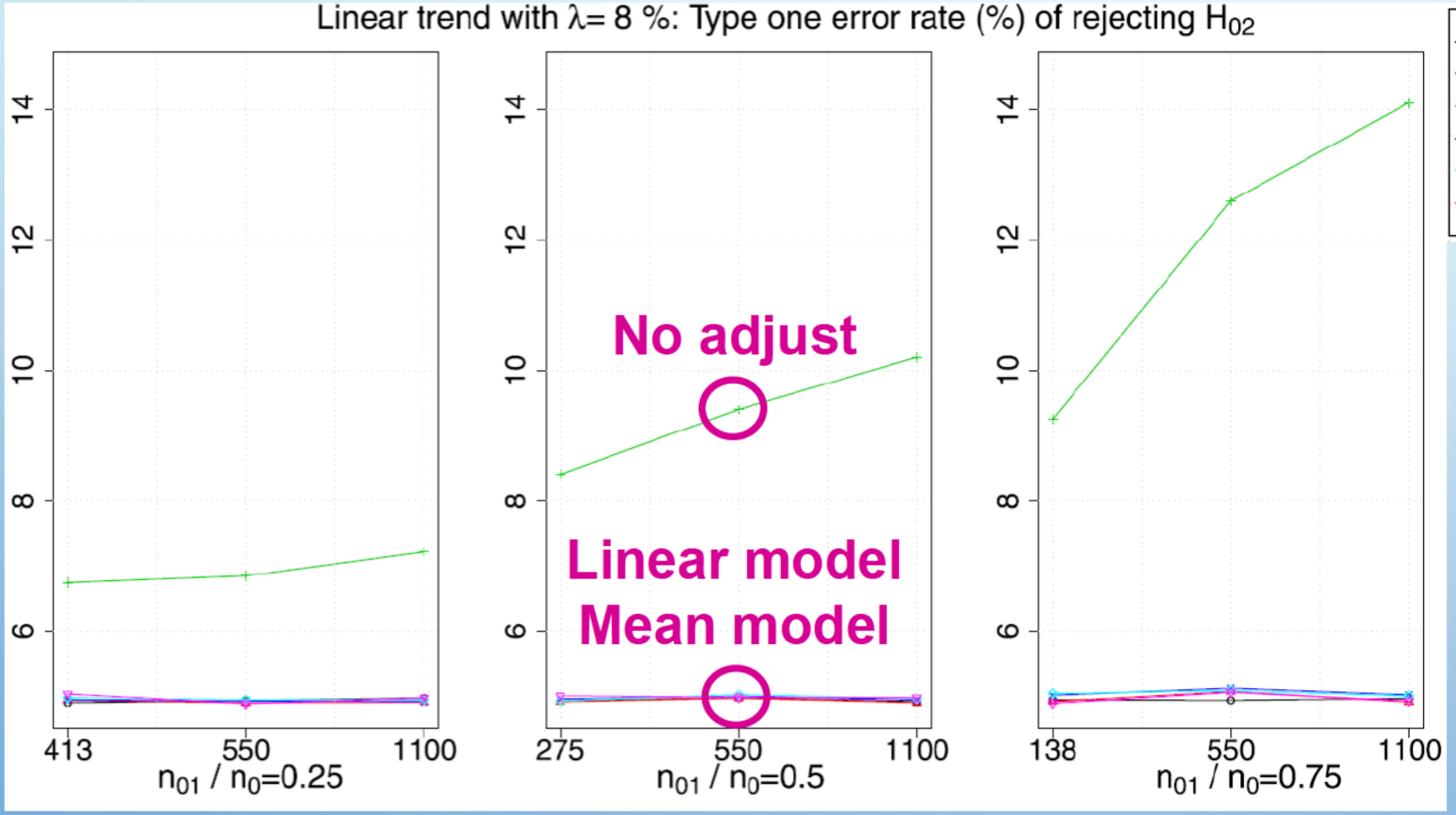
No adjust **Linear model** **Two-arm Mean model**

Power of No adjust, Linear and Mean model for Linear Trend



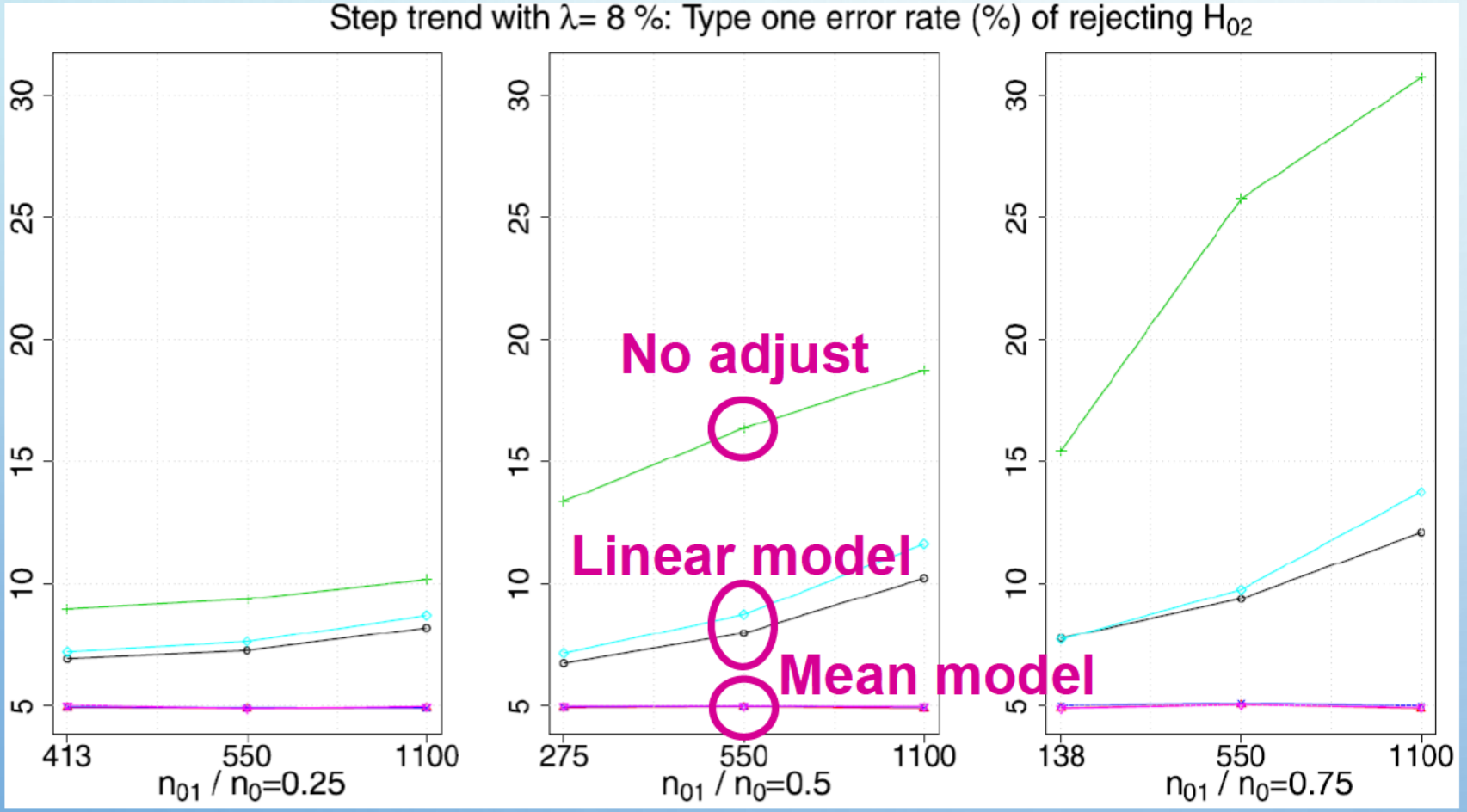
- Utilize all treat. arms
  - $M_{a1}: X_{jk} = \beta_k + \gamma \cdot j + \epsilon_j$
  - $M_{a2}: X_{jk} = \beta_k + \nu \cdot I(j \in stage\ 2) + \epsilon_j$
- Use only 2 arms
  - $M_{b1}: X_{jk} = \beta_k + \gamma \cdot j + \epsilon_j$
  - $M_{b2}: X_{jk} = \beta_k + \nu \cdot I(j \in stage\ 2) + \epsilon_j$

Type I error rate under Linear Trend



- Utilize all treat. arms
  - $M_{a1}: X_{jk} = \beta_k + \gamma \cdot j + \epsilon_j$
  - $M_{a2}: X_{jk} = \beta_k + \nu \cdot I(j \in stage\ 2) + \epsilon_j$
- Use only 2 arms
  - $M_{b1}: X_{jk} = \beta_k + \gamma \cdot j + \epsilon_j$
  - $M_{b2}: X_{jk} = \beta_k + \nu \cdot I(j \in stage\ 2) + \epsilon_j$

Type I error rate under **Step** Trend



- Utilize all treat. arms
  - $M_{a1}: X_{jk} = \beta_k + \gamma \cdot j + \epsilon_j$
  - $M_{a2}: X_{jk} = \beta_k + \nu \cdot I(j \in stage\ 2) + \epsilon_j$
- Use only 2 arms
  - $M_{b1}: X_{jk} = \beta_k + \gamma \cdot j + \epsilon_j$
  - $M_{b2}: X_{jk} = \beta_k + \nu \cdot I(j \in stage\ 2) + \epsilon_j$

Modeling time drifts in the control group

- Including time as a continuous or categorical covariate
  - Control the Type I error rate if the model assumptions are correct (Lee and Wason, 2020)
  - Can **NOT control rigorously**, but can **suppress considerable** inflation in case of no adjustment
- Comparison including NC control depends on
  - **Position** of analysis result
  - **Absolute size** and **monotonicity** of time drift

Modeling time drifts in the control group

- Other issues
  - Need of controlling **measured confounders**, time drift deals with **unmeasured part** ?
  - Especially when NC control **size is large**, dynamic borrowing may be needed such as **Hybrid Control Approach**

Thank you for your attention !