

# The Bayesian Time Machine: Accounting for temporal drift in multi-arm platform trials

Ben Saville, Ph.D.

Director & Senior Statistical Scientist, Berry Consultants

Adjunct Assistant Professor, Vanderbilt University



# Outline

1. Introduction to Platform Trials
2. Temporal Drift
  - Motivating example
  - Methods
  - Simulations
3. PRINCIPLE trial
  - Temporal adjustment in COVID

REVIEW ARTICLE

THE CHANGING FACE OF CLINICAL TRIALS

Jeffrey M. Drazen, M.D., David P. Harrington, Ph.D., John J.V. McMurray, M.D., James H. Ware, Ph.D.,  
and Janet Woodcock, M.D., *Editors*

## Master Protocols to Study Multiple Therapies, Multiple Diseases, or Both

Janet Woodcock, M.D., and Lisa M. LaVange, Ph.D.

From the Center for Drug Evaluation and Research, Food and Drug Administration, Silver Spring, MD. Address reprint requests to Dr. LaVange at the Office of Biostatistics, Office of Translational Sciences, Center for Drug Evaluation and Research, Food and Drug Administration, 10903 New Hampshire Blvd., Silver Spring, MD 20993, or at lisa.lavange@fda.hhs.gov.

N Engl J Med 2017;377:62-70.

DOI: 10.1056/NEJMra1510062

Copyright © 2017 Massachusetts Medical Society.

**H**IGH-QUALITY EVIDENCE IS WHAT WE USE TO GUIDE MEDICAL PRACTICE. The standard approach to generating this evidence — a series of clinical trials, each investigating one or two interventions in a single disease — has become ever more expensive and challenging to execute. As a result, important clinical questions go unanswered. The conduct of “precision medicine” trials to evaluate targeted therapies creates challenges in recruiting patients with rare genetic subtypes of a disease. There is also increasing interest in performing mechanism-based trials in which eligibility is based on criteria other than traditional disease definitions. The common denominator is a need to answer more questions more efficiently and in less time.

A methodologic innovation responsive to this need involves coordinated efforts to evaluate more than one or two treatments in more than one patient type or disease within the same overall trial structure.<sup>1-4</sup> Such efforts are referred to as master protocols, defined as one overarching protocol designed to answer multiple questions. Master protocols may involve one or more interventions in multiple diseases or a single disease, as defined by current disease classification, with multiple interventions, each targeting a particular biomarker-defined population or disease subtype. Included under this broad definition of a master protocol are three distinct entities: umbrella, basket, and platform trials (Table 1 and Figs. 1 and 2). All constitute a

# Adaptive Platform Trials

- Master Protocol
- Focus is on the Disease
  - “What is the best treatment for a unique patient with this disease?
- Typical Innovations
  - Staggered entry of interventions
  - Graduation/Removal, “Perpetual” trials
  - Response Adaptive Randomization (RAR)
  - Patient heterogeneity (hierarchical modeling)
  - Combination treatments
  - Statistical Modeling

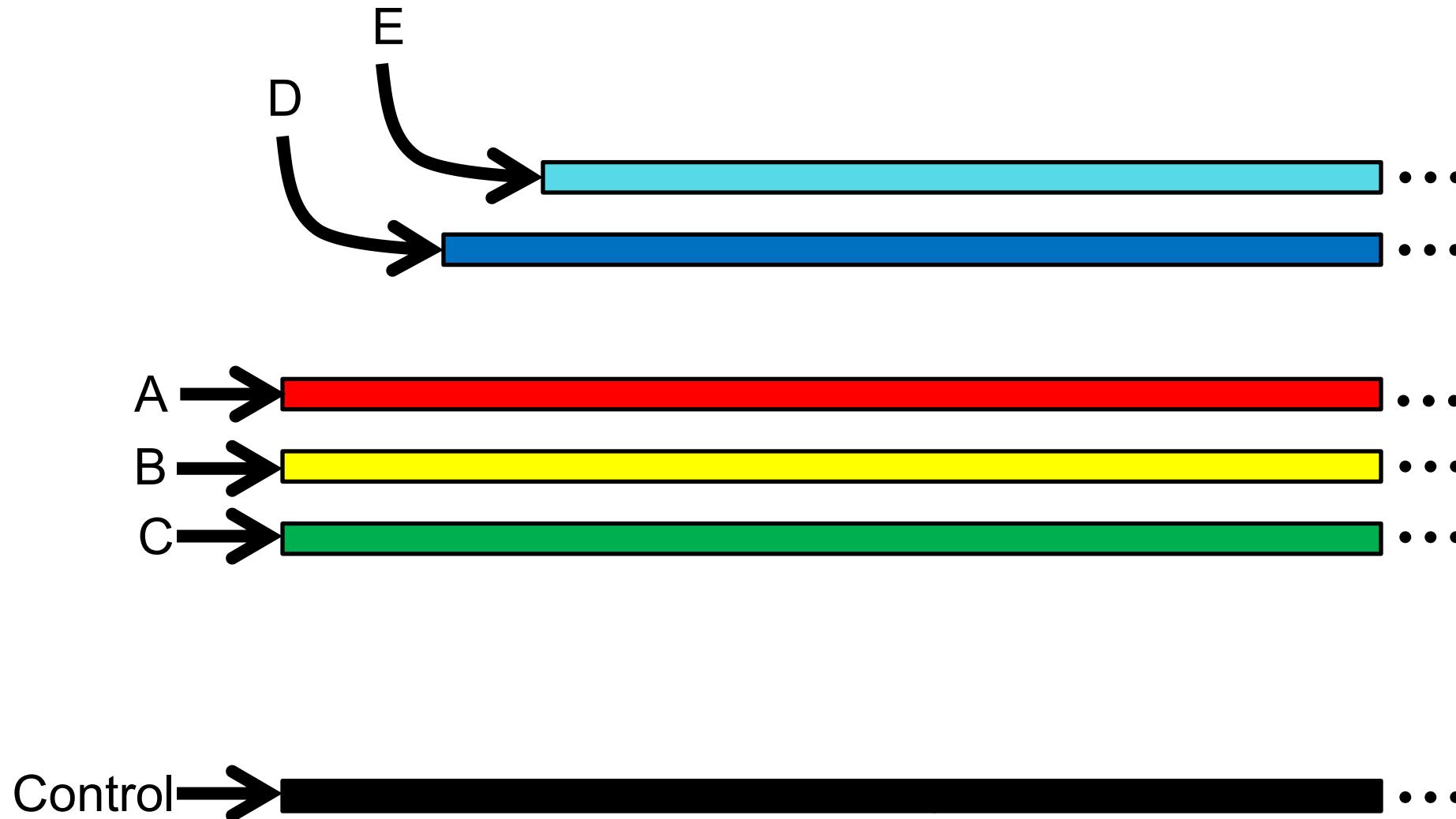
# Potential Features of a Platform Trial

---



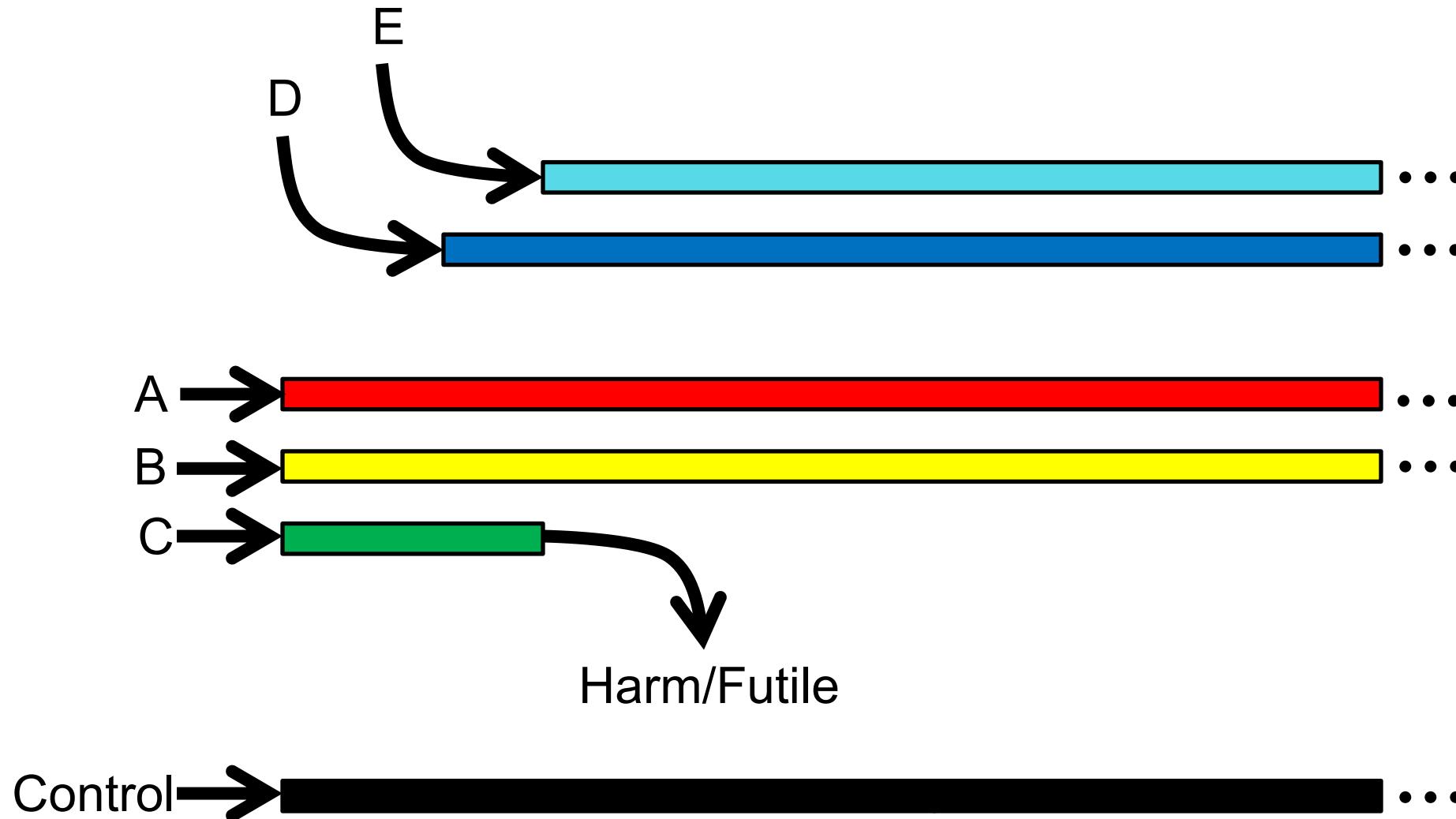
# Potential Features of a Platform Trial

---



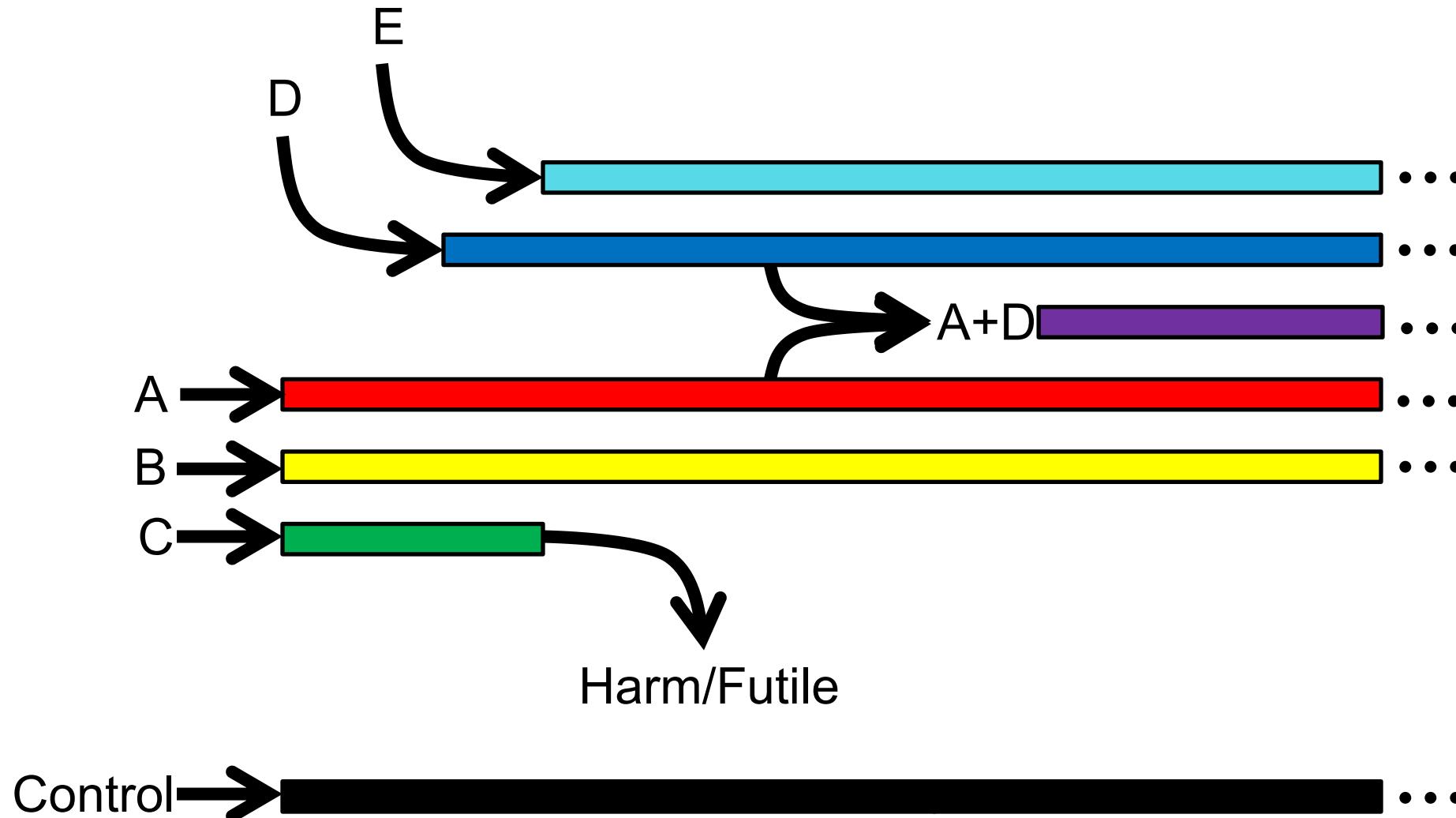
# Potential Features of a Platform Trial

---



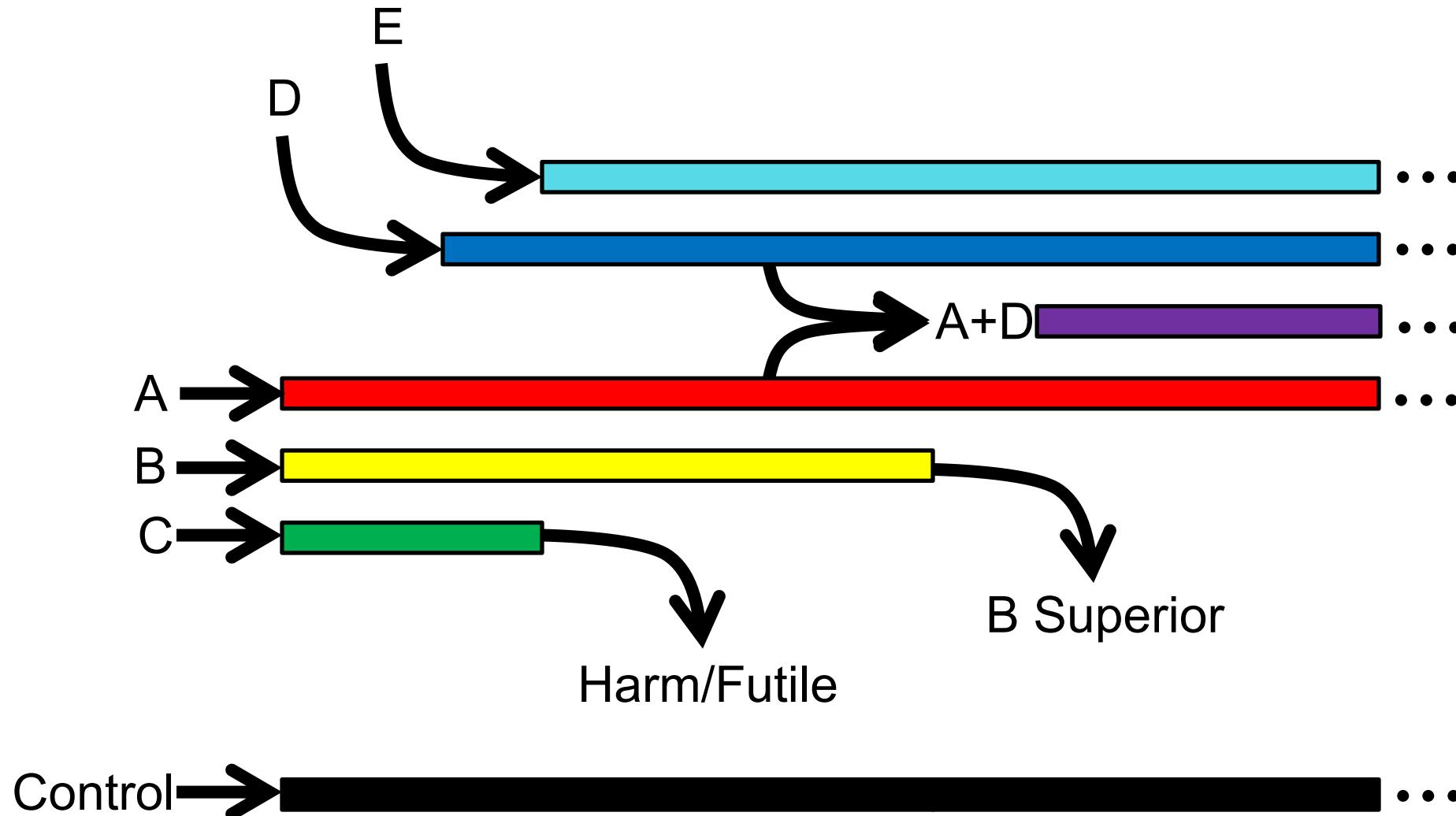
# Potential Features of a Platform Trial

---

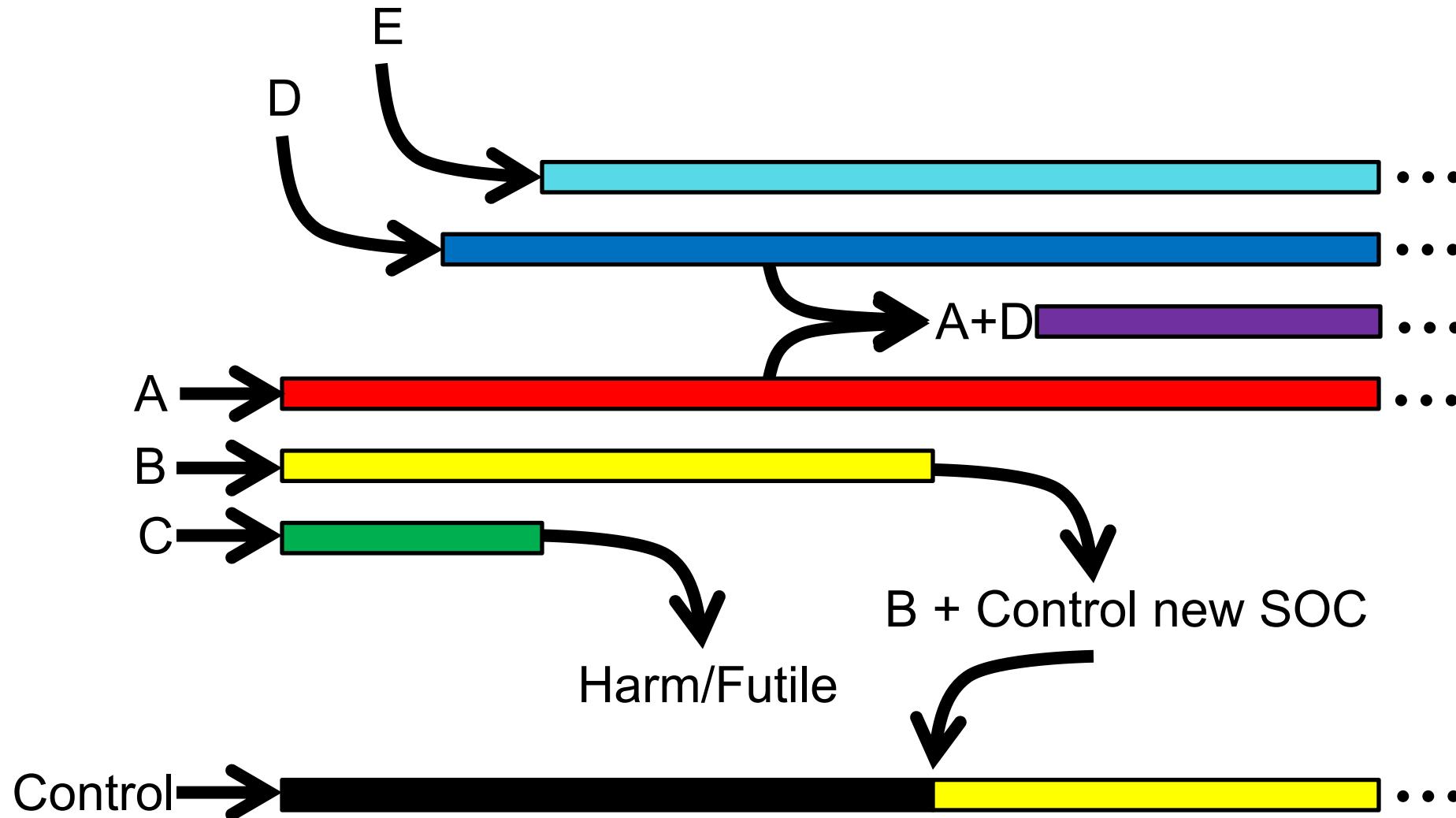


# Potential Features of a Platform Trial

---



# Potential Features of a Platform Trial



# Platform Trials and Temporal Drift

- Platform trials typically span a longer time period than traditional clinical trials
- Typically have staggered entry of interventions
  - Randomization ratios may be changing (RAR)
- Regulators and community have expressed concern over impact of “temporal drift” on platform trials
  - Changes in population, SOC, disease (COVID strain), etc.
- How does one leverage all data on control arms from earlier in trial?

# The Bayesian Time Machine: Accounting for temporal drift in multi-arm platform trials

*Clinical Trials*

1–12

© The Author(s) 2022

Article reuse guidelines:

[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)

DOI: [10.1177/17407745221112013](https://doi.org/10.1177/17407745221112013)

[journals.sagepub.com/home/ctj](https://journals.sagepub.com/home/ctj)



**Benjamin R Saville<sup>1,2</sup> , Donald A Berry<sup>1,3</sup>, Nicholas S Berry<sup>1</sup>,**  
**Kert Viele<sup>1,4</sup> and Scott M Berry<sup>1,5</sup>**

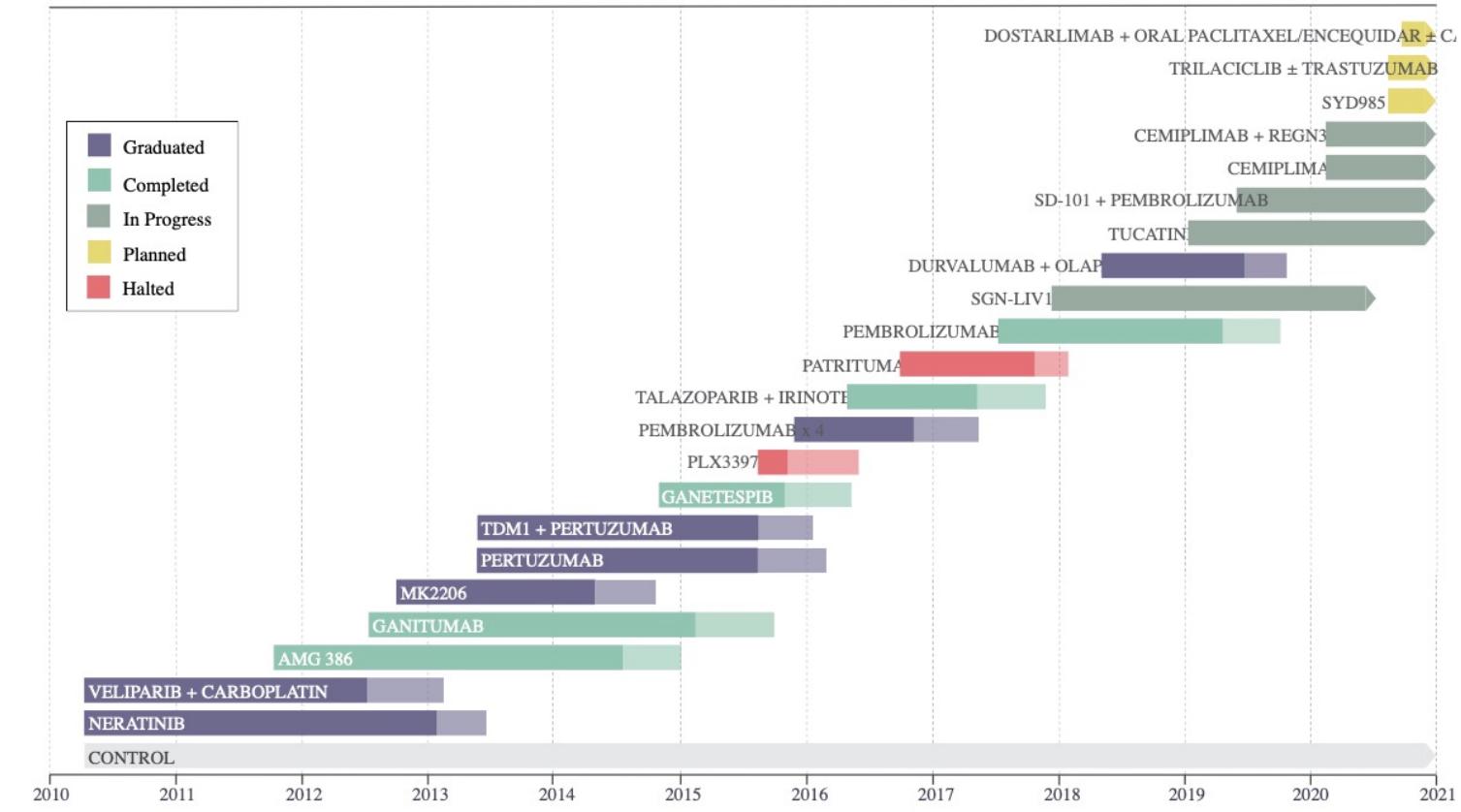
## Abstract

**Background:** Multi-arm platform trials investigate multiple agents simultaneously, typically with staggered entry and exit of experimental treatment arms versus a shared control arm. In such settings, there is considerable debate whether to limit analyses for a treatment arm to concurrent randomized control subjects or to allow comparisons to both concurrent and non-concurrent (pooled) control subjects. The potential bias from temporal drift over time is at the core of this debate.

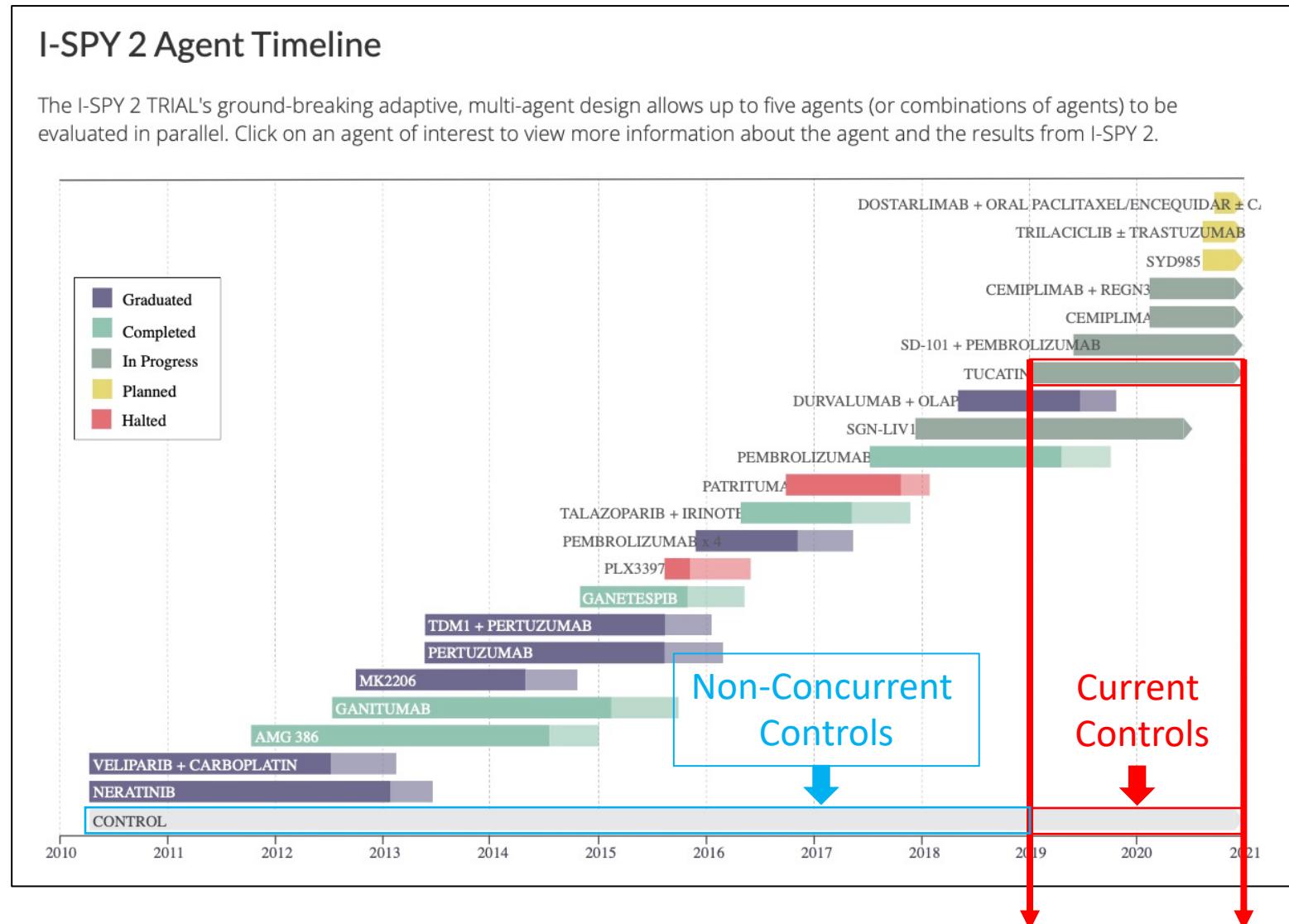
# Platform Trial & Non-Concurrent Controls

## I-SPY 2 Agent Timeline

The I-SPY 2 TRIAL's ground-breaking adaptive, multi-agent design allows up to five agents (or combinations of agents) to be evaluated in parallel. Click on an agent of interest to view more information about the agent and the results from I-SPY 2.



# Platform Trial & Non-Concurrent Controls

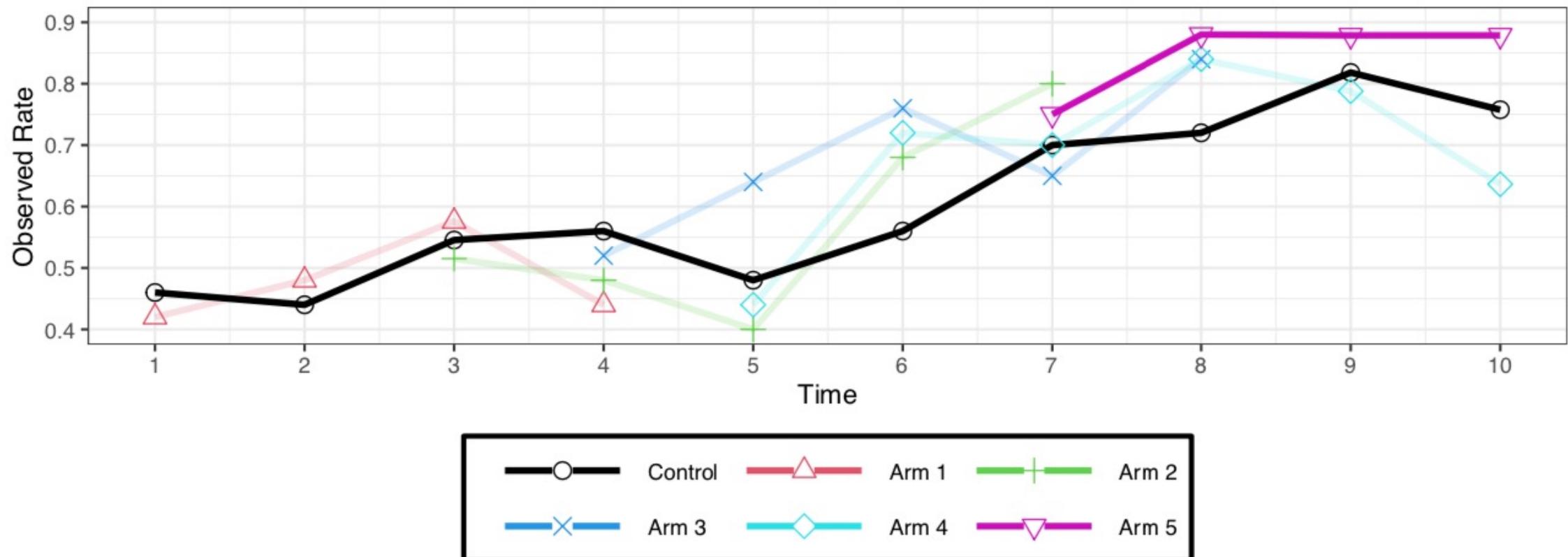


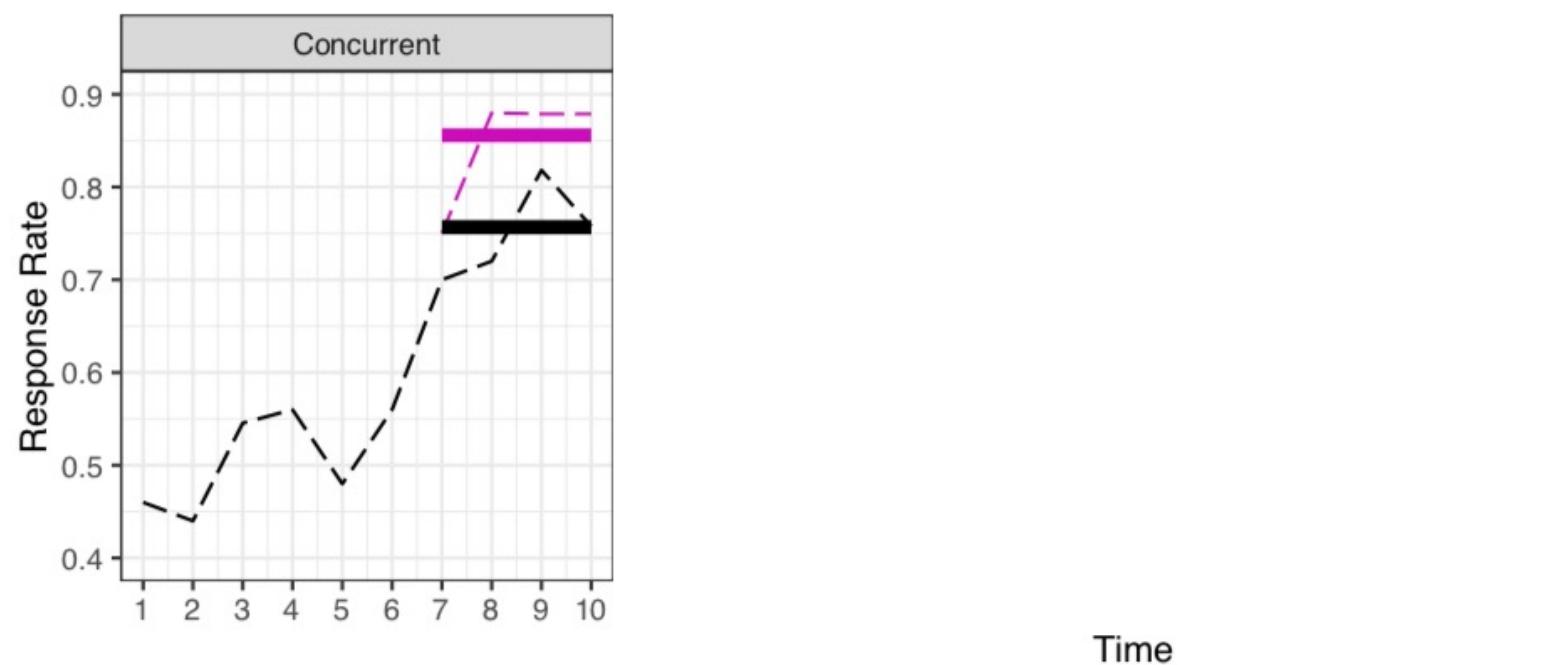
# Historical Controls

- “Historical Controls” is a generally a *pejorative* term (to some)
- Non-Concurrent controls in platform trials are randomized
  - in the same protocol
  - same inclusion/exclusion
  - same visits
  - same procedures
  - same data quality
  - overlapping treatment arms
- The only difference is **TIME!!**

# How to Model?

- What is the best way to estimate effect of Arm 5?

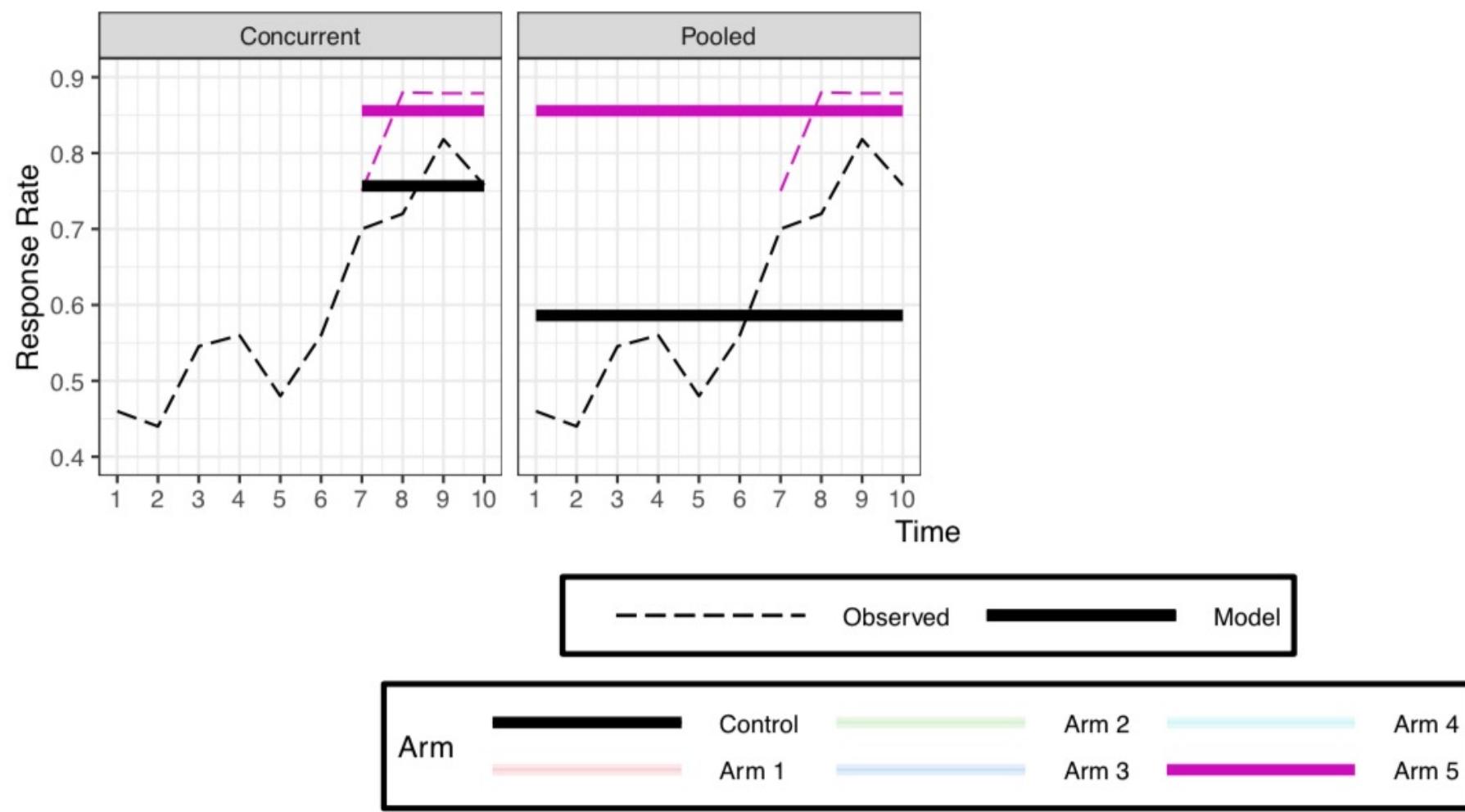




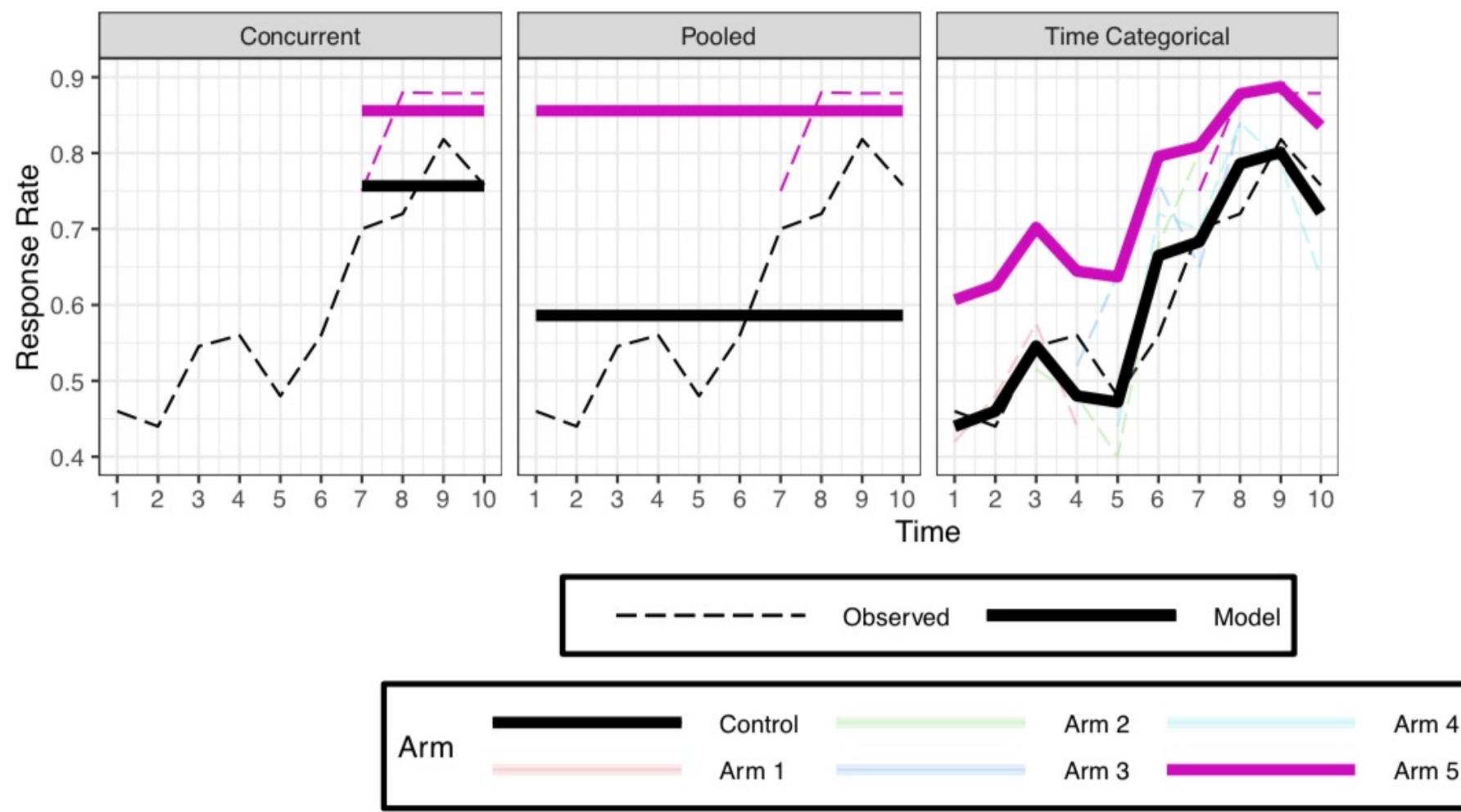
----- Observed    ━━━ Model

Arm    ━━━ Control    ━━━━ Arm 2    ━━━━ Arm 4  
           ━━━━ Arm 1    ━━━━ Arm 3    ━━━━ Arm 5

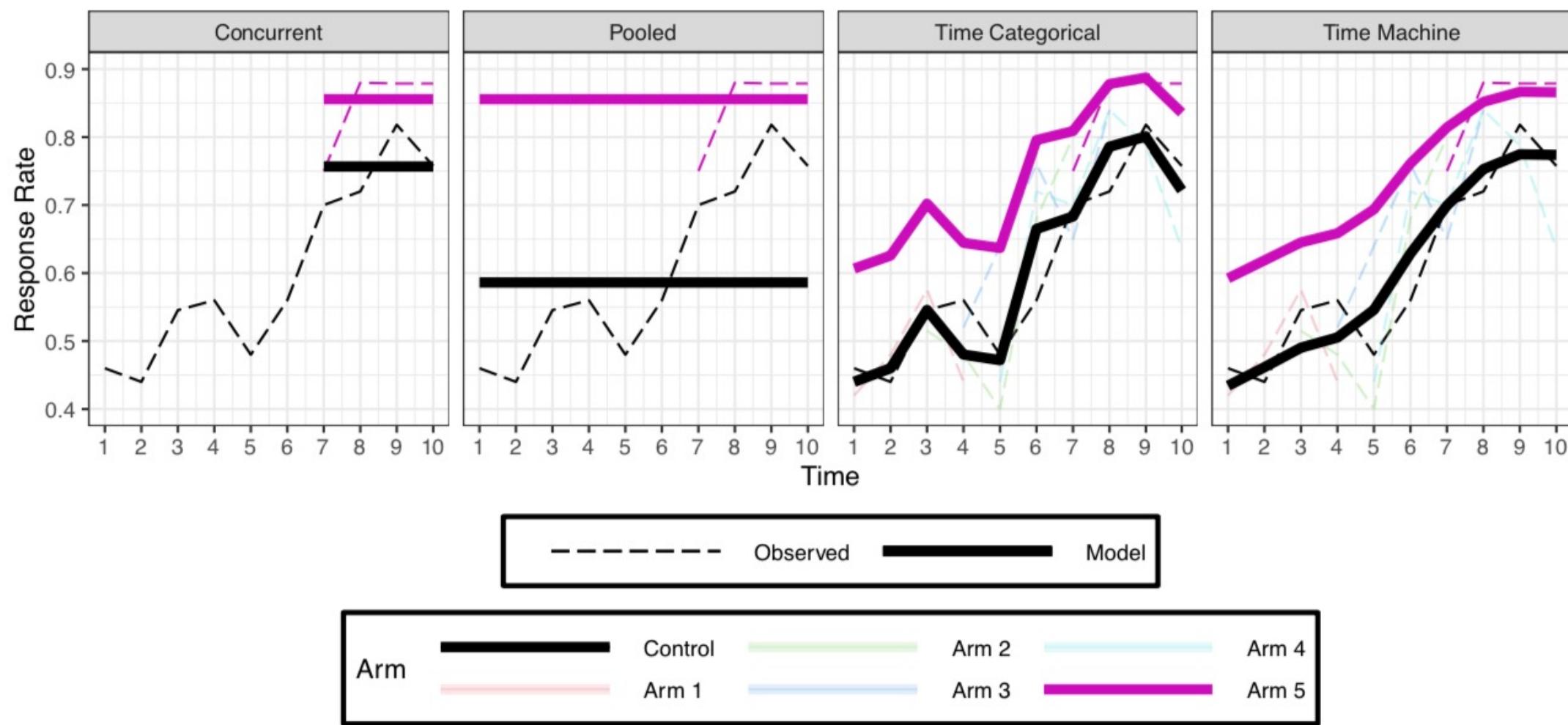
Method	Log Odds Ratio	Standard Error	Odds Ratio (95% CI)	One-sided P-value	Posterior Prob
1) Concurrent Controls	0.646	0.349	1.91 (0.96,3.78)	0.032	-



Method	Log Odds Ratio	Standard Error	Odds Ratio (95% CI)	One-sided P-value	Posterior Prob
1) Concurrent Controls	0.646	0.349	1.91 (0.96,3.78)	0.032	-
2) Pooled Controls	1.433	0.293	4.19 (2.36,7.45)	0	-



Method	Log Odds Ratio	Standard Error	Odds Ratio (95% CI)	One-sided P-value	Posterior Prob
1) Concurrent Controls	0.646	0.349	1.91 (0.96,3.78)	0.032	-
2) Pooled Controls	1.433	0.293	4.19 (2.36,7.45)	0	-
3) Time Categorical	0.674	0.328	1.96 (1.03,3.73)	0.02	-



Method	Log Odds Ratio	Standard Error	Odds Ratio (95% CI)	One-sided P-value	Posterior Prob
1) Concurrent Controls	0.646	0.349	1.91 (0.96,3.78)	0.032	-
2) Pooled Controls	1.433	0.293	4.19 (2.36,7.45)	0	-
3) Time Categorical	0.674	0.328	1.96 (1.03,3.73)	0.02	-
4) Time Machine	0.653	0.315	1.92 (1.04,3.56)	-	0.981

# Bayesian Time Machine

- Second-order normal dynamic linear model (NDLM)

$$\begin{aligned}E(Y_i) &= \mu_i \\g(\mu_i) &= \gamma + \theta_{j(i)} + \mathbf{x}'_i \boldsymbol{\beta} + \alpha_{t(i)}\end{aligned}$$

- $Y_i$ : response for subject  $i$
- $\gamma$ : model intercept
- $\theta_{j(i)}$ : increment in linear predictor for treatment arm  $j$
- $\alpha_{t(i)}$ : increment in linear predictor for time interval  $t$
- $\mathbf{x}$ : vector of covariates with parameters  $\boldsymbol{\beta}$
- Non-informative priors for  $\gamma, \theta_{j(i)}, \boldsymbol{\beta}$

# Temporal Drift

- Divide time since start of trial into  $T$  buckets
  - Count backwards from most recent time interval ( $t = 1$ ) to beginning of trial ( $t + 1, t + 2, \dots, T$ )
$$\alpha_1 = 0 \text{ for identifiability}$$
$$\alpha_2 \sim N(0, 1/\tau)$$
$$\alpha_t \sim N(2\alpha_{t-1} - \alpha_{t-2}, 1/\tau), t \geq 3$$
- $\tau$ : Drift parameter that determines amount of “smoothing”

# Dynamic Smoothing

- Use observed data to determine degree of smoothing
- Hyperprior for  $\tau$

$$\tau \sim \text{Gamma}(a, b)$$

- Weight of  $2a$  intervals of data centered at  $a/b$
- Centering  $\tau$  at large values with large variance allows flexible and dynamic smoothing
  - Estimated drift determined by observed data

# Simulation Study

## Objective

- Simulate a platform trial **with** and **without** temporal trends, and compare different analysis strategies

## Assumptions/Setting

- Binary endpoint
- 5 treatment arms vs. control (focus on Arm 5)
- Staggered entry and exit of treatment arms
- 10 distinct time intervals
- 100 patients randomized within each time interval
- No lag between randomization and primary outcome measurement

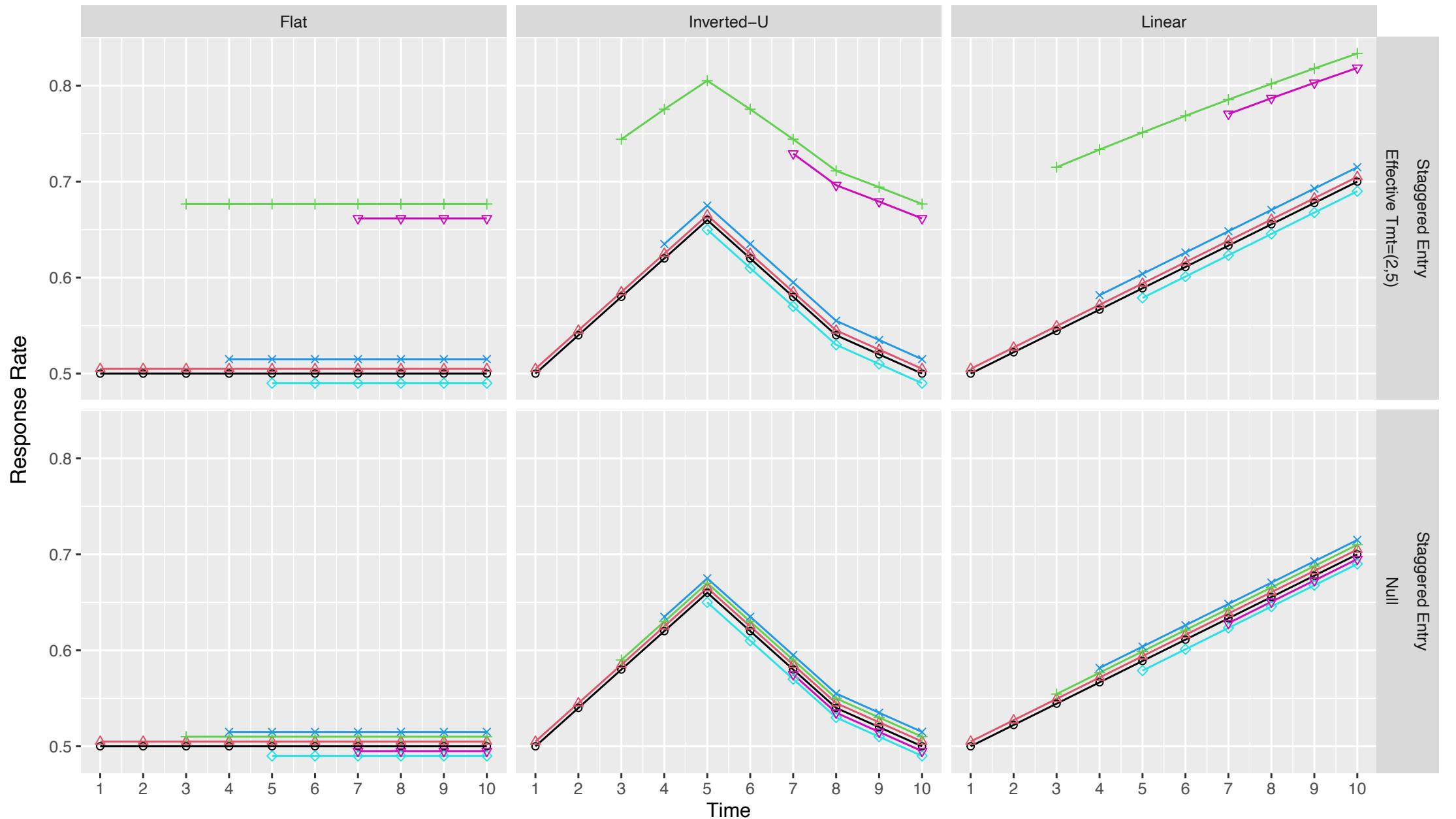
# Data Assumptions (Base)

- Control response rate: 50% at start of trial
  1. Rate constant
  2. Rate increases to a peak; then decreases (inverted-U)
  3. Rate has linear increase
- Treatment response rates
  1. No benefit vs. control: Odds ratio = 1.0
  2. Superior vs. control: Odds ratio = 2.0
- Staggered entry (& exit) of treatment arms

# Number of Subjects by time (Base)

**Table I.** Scenarios: Number of subjects by arm and time.

Staggered	Treatment	Time 1	Time 2	Time 3	Time 4	Time 5	Time 6	Time 7	Time 8	Time 9	Time 10
Entry	Control	50	50	33	25	20	20	17	17	17	17
	Arm 1	50	50	33	25	20	20	17	17	17	17
	Arm 2			33	25	20	20	17	17	17	17
	Arm 3				25	20	20	17	17	17	17
	Arm 4					20	20	17	17	17	17
	Arm 5						17	17	17	17	17



# Analysis Methods for Arm 5 vs. Control

1. Concurrent Controls Analysis
  - Logistic regression for Arm 5 vs. concurrent controls
2. Pooled Controls Analysis
  - Logistic regression for Arm 5 vs. pooled controls
3. Time Categorical Analysis
  - Logistic regression for Arm 5 vs. all controls
  - Covariate adjustment for time (categorical bins)
  - Includes all treatment arms
4. Bayesian Time Machine
  - Bayesian logistic regression for Arm 5 vs. all controls
  - Adjustment for time via Bayesian smoothing
  - Includes all treatment arms

# Analysis Methods: Assumptions

1. Concurrent Controls Analysis
  - Assumes non-concurrent controls have no useful information
2. Pooled Controls Analysis
  - Assumes control response rate is constant across time
3. Time Categorical Analysis
  - Assumes independent bins for control response across time
4. Bayesian Time Machine
  - Assumes smoothing control response across time is appropriate
  - $\tau^2 \sim \text{Gamma}(a = 0.1, b = 0.01)$

All strategies assume treatment effects are constant across time

# Metrics

1. Bias and standard error (SE) of treatment effects
2. Mean square error (MSE) ratio of each method compared to Bayesian Time Machine
3. Statistical Power
  - $H_0$ : log-odds of Arm 5 vs. control  $\leq 0$
  - $H_1$ : log-odds of Arm 5 vs. control  $> 0$
  - Use one-sided alpha = 0.025 for p-values; Bayes posterior prob 0.975

# Results

**Table 2.** Treatment Arm 5 vs. Control

			Avg Bias log(OR)				Average SE				MSE Ratio vs. TimeMachine (smaller is better)				Statistical Power			
Stag	Eff	Drift	ConC	PoolC	Time Cat	Time Mach	ConC	PoolC	Time Cat	Time Mach	ConC	PoolC	Time Cat	Time Mach	ConC	PoolC	Time Cat	Time Mach
E	Null	Flat	0.00	0.00	0.00	0.00	0.35	0.27	0.30	0.29	1.44	0.91	1.08	1	0.022	0.023	0.026	0.024
		Lin	0.00	0.41	0.00	0.02	0.37	0.29	0.31	0.31	1.44	2.64	1.08	1	0.023	0.274	0.024	0.027
		InvU	0.00	-0.10	0.00	-0.02	0.35	0.27	0.30	0.29	1.42	0.99	1.07	1	0.028	0.010	0.026	0.022
2,5		Flat	0.01	0.01	0.01	-0.01	0.36	0.29	0.31	0.30	1.42	0.93	1.08	1	0.493	0.693	0.631	0.631
		Lin	0.02	0.43	0.03	0.03	0.41	0.33	0.36	0.35	1.37	2.43	1.10	1	0.409	0.954	0.524	0.565
		InvU	0.01	-0.08	0.02	-0.01	0.36	0.29	0.32	0.31	1.39	0.99	1.08	1	0.478	0.536	0.613	0.598

# Results

**Table 2.** Treatment Arm 5 vs. Control

Stage	Eff	Drift
E	Null	Flat Lin InvU
2,5	Flat Lin InvU	

# Results

**Table 2.** Treatment Arm 5 vs. Control

			Avg Bias log(OR)			
Stag	Eff	Drift	ConC	PoolC	Time Cat	Time Mach
			0.00	0.00	0.00	0.00
E	Null	Flat	0.00	0.41	0.00	0.02
		Lin	0.00	-0.10	0.00	-0.02
		InvU	0.01	0.01	0.01	-0.01
2,5		Flat	0.02	0.43	0.03	0.03
		Lin	0.01	-0.08	0.02	-0.01
		InvU				

# Results

**Table 2.** Treatment Arm 5 vs. Control

			Avg Bias log(OR)				Average SE			
Stag	Eff	Drift	ConC	PoolC	Time Cat	Time Mach	ConC	PoolC	Time Cat	Time Mach
E	Null	Flat	0.00	0.00	0.00	0.00	0.35	0.27	0.30	0.29
		Lin	0.00	0.41	0.00	0.02	0.37	0.29	0.31	0.31
		InvU	0.00	-0.10	0.00	-0.02	0.35	0.27	0.30	0.29
2,5		Flat	0.01	0.01	0.01	-0.01	0.36	0.29	0.31	0.30
		Lin	0.02	0.43	0.03	0.03	0.41	0.33	0.36	0.35
		InvU	0.01	-0.08	0.02	-0.01	0.36	0.29	0.32	0.31

# Results

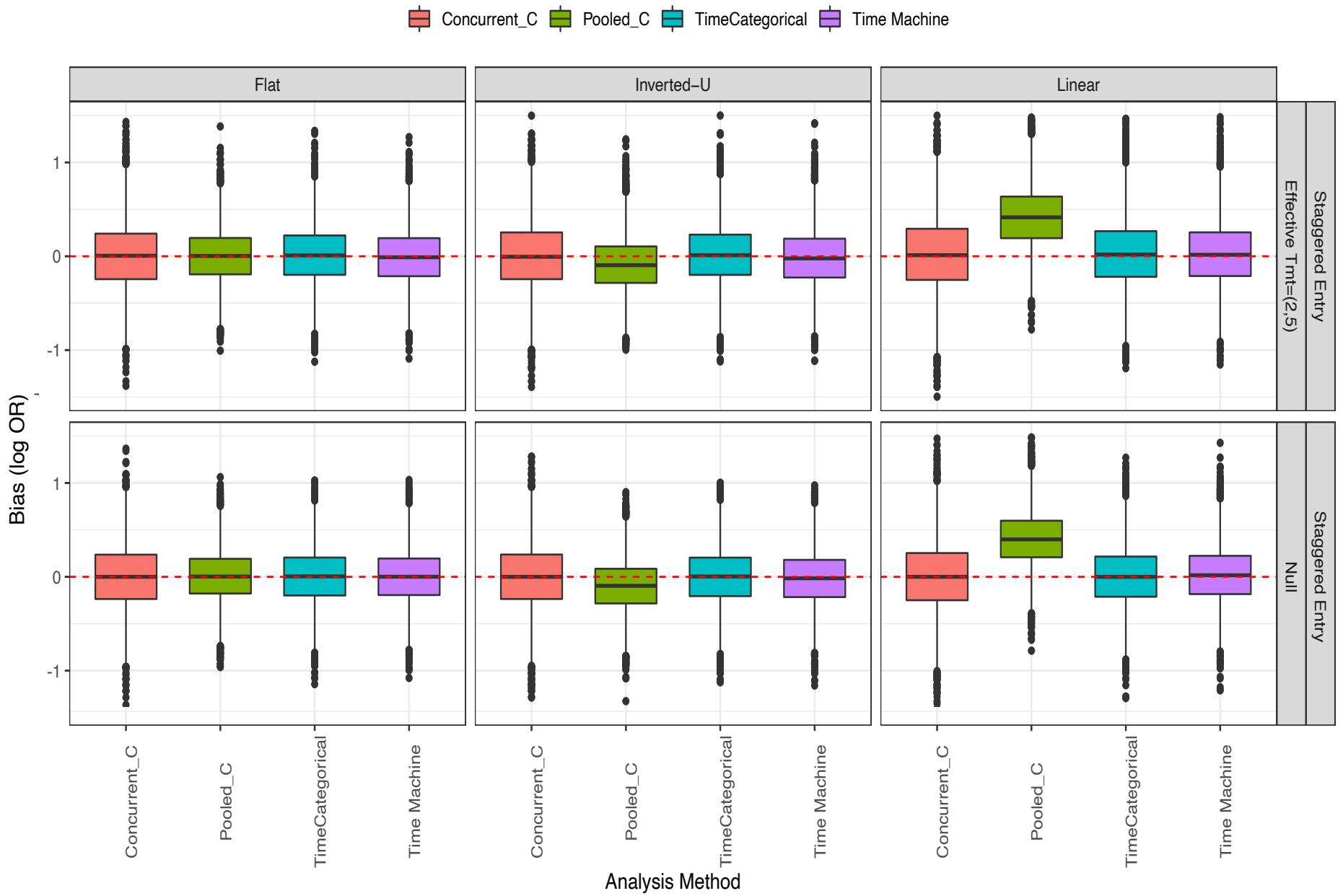
**Table 2.** Treatment Arm 5 vs. Control

			Avg Bias log(OR)				Average SE				MSE Ratio vs. TimeMachine (smaller is better)			
Stag	Eff	Drift	ConC	PoolC	Time Cat	Time Mach	ConC	PoolC	Time Cat	Time Mach	ConC	PoolC	Time Cat	Time Mach
E	Null	Flat	0.00	0.00	0.00	0.00	0.35	0.27	0.30	0.29	1.44	0.91	1.08	1
		Lin	0.00	0.41	0.00	0.02	0.37	0.29	0.31	0.31	1.44	2.64	1.08	1
		InvU	0.00	-0.10	0.00	-0.02	0.35	0.27	0.30	0.29	1.42	0.99	1.07	1
2,5		Flat	0.01	0.01	0.01	-0.01	0.36	0.29	0.31	0.30	1.42	0.93	1.08	1
		Lin	0.02	0.43	0.03	0.03	0.41	0.33	0.36	0.35	1.37	2.43	1.10	1
		InvU	0.01	-0.08	0.02	-0.01	0.36	0.29	0.32	0.31	1.39	0.99	1.08	1

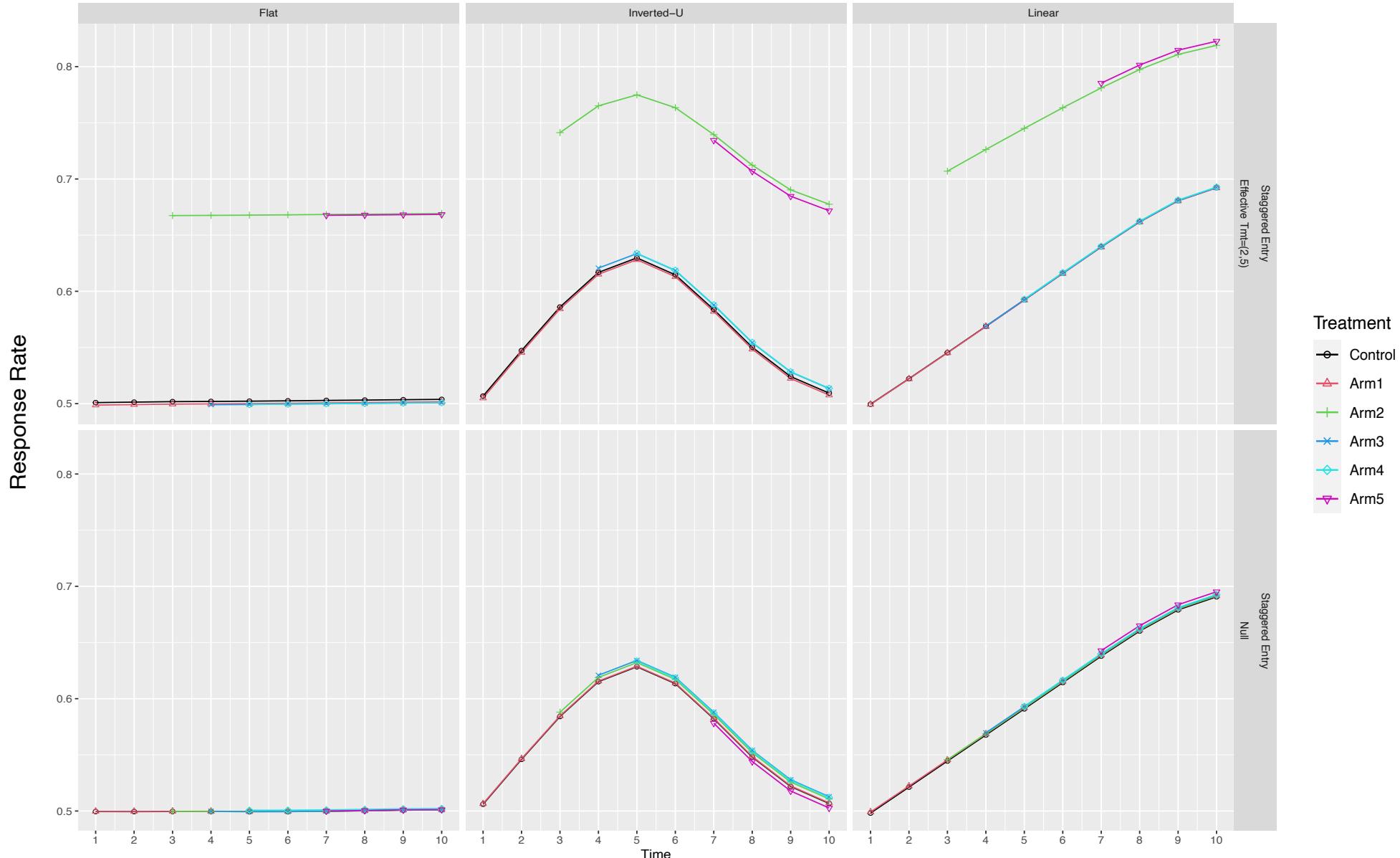
# Results

**Table 2.** Treatment Arm 5 vs. Control

			Avg Bias log(OR)				Average SE				MSE Ratio vs. TimeMachine (smaller is better)				Statistical Power			
Stag	Eff	Drift	ConC	PoolC	Time Cat	Time Mach	ConC	PoolC	Time Cat	Time Mach	ConC	PoolC	Time Cat	Time Mach	ConC	PoolC	Time Cat	Time Mach
E	Null	Flat	0.00	0.00	0.00	0.00	0.35	0.27	0.30	0.29	1.44	0.91	1.08	1	0.022	0.023	0.026	0.024
		Lin	0.00	0.41	0.00	0.02	0.37	0.29	0.31	0.31	1.44	2.64	1.08	1	0.023	0.274	0.024	0.027
		InvU	0.00	-0.10	0.00	-0.02	0.35	0.27	0.30	0.29	1.42	0.99	1.07	1	0.028	0.010	0.026	0.022
2,5		Flat	0.01	0.01	0.01	-0.01	0.36	0.29	0.31	0.30	1.42	0.93	1.08	1	0.493	0.693	0.631	0.631
		Lin	0.02	0.43	0.03	0.03	0.41	0.33	0.36	0.35	1.37	2.43	1.10	1	0.409	0.954	0.524	0.565
		InvU	0.01	-0.08	0.02	-0.01	0.36	0.29	0.32	0.31	1.39	0.99	1.08	1	0.478	0.536	0.613	0.598



# Average Estimates: Time Machine



# Simulation Results

- Time Machine inferential improvements
  - Power, MSE: 20% improvement vs concurrent controls
  - Improvement applies to EACH arm in the platform!
- Time Machine nearly unbiased
  - Smoothed estimates across time more plausible
- Time Categorical also outperforms pooled and concurrent controls
  - Adjusting for time is superior to either extreme



**The PRINCIPLE Adaptive Platform Trial  
for Community Treatment of COVID-19:  
Innovation in Trial Design and Delivery**



# PRINCIPLE: COVID-19 in Primary Care

- Most people with COVID-19 are managed in the community
  - Community treatments may have the widest reach and impact
- PRINCIPLE objective: Evaluate whether re-purposed drugs can make a difference with early intervention
- Needed a rapidly initiated trial with adaptive features
  - Ability to evaluate treatments quickly (early superiority/futility)
  - Flexibility to add treatments
- Urgency: First patient randomized < 3 weeks from initial contact with Oxford collaborators!

# PRINCIPLE: COVID-19 in Primary Care

Participants:

- Presenting **in primary care** within 14 days since onset of cough and/or fever during time of prevalent COVID-19 infections

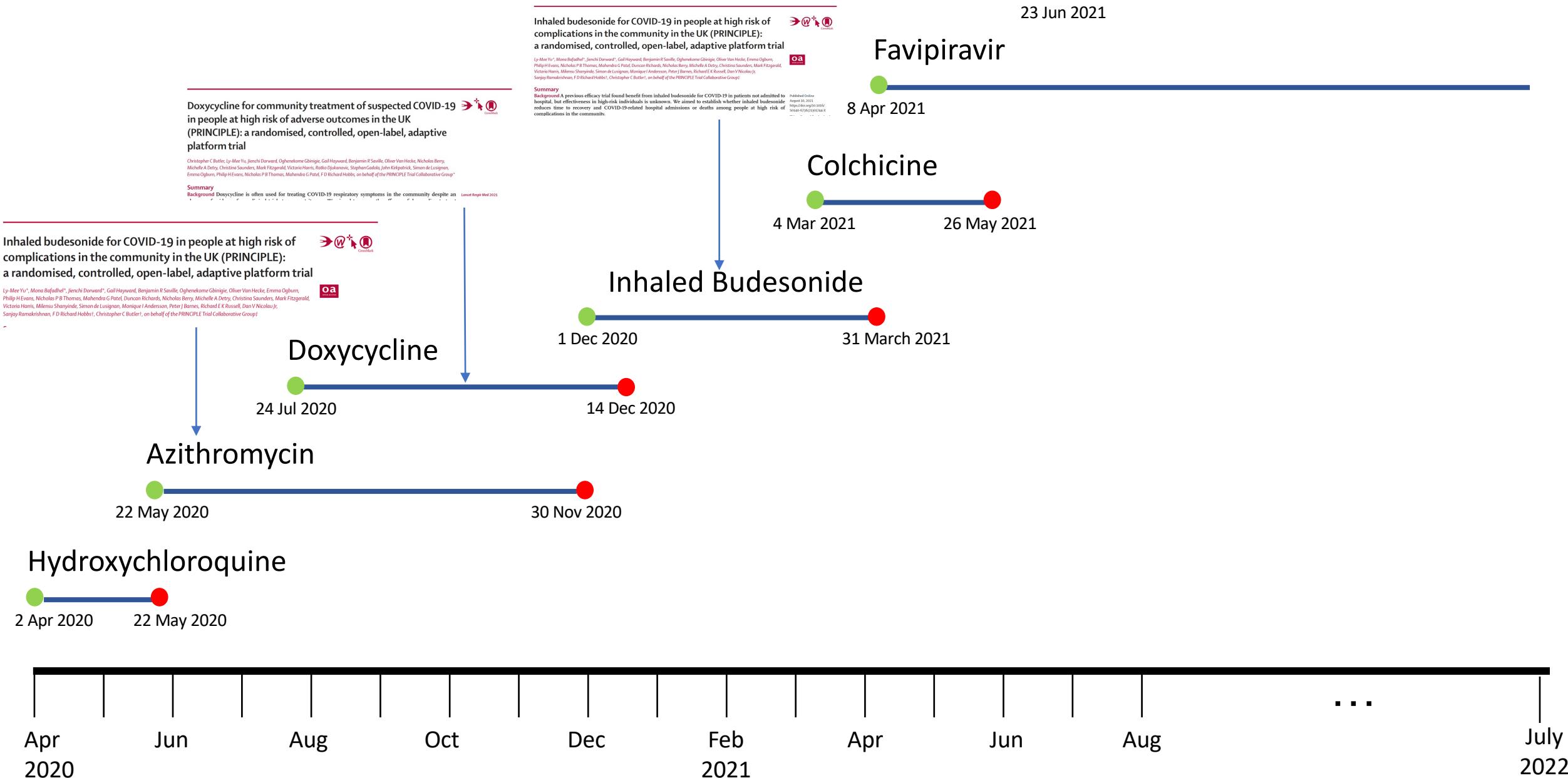
Randomized among multiple interventions or Usual Care

- Frequent interim analyses
- Response adaptive randomization

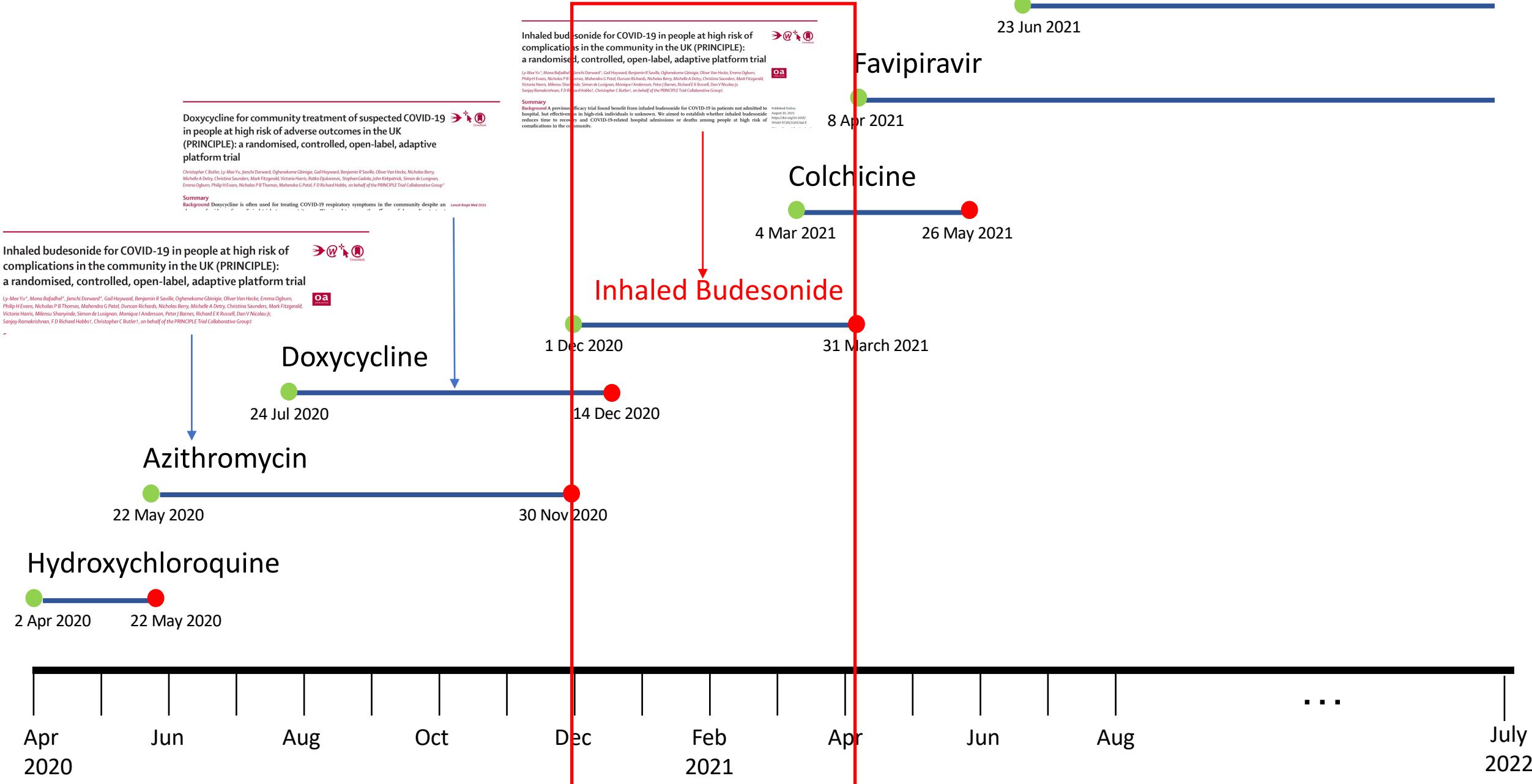
Primary endpoints:

1. Time to self-reported recovery within 28 days
2. Hospitalization/Death (binary) within 28 days

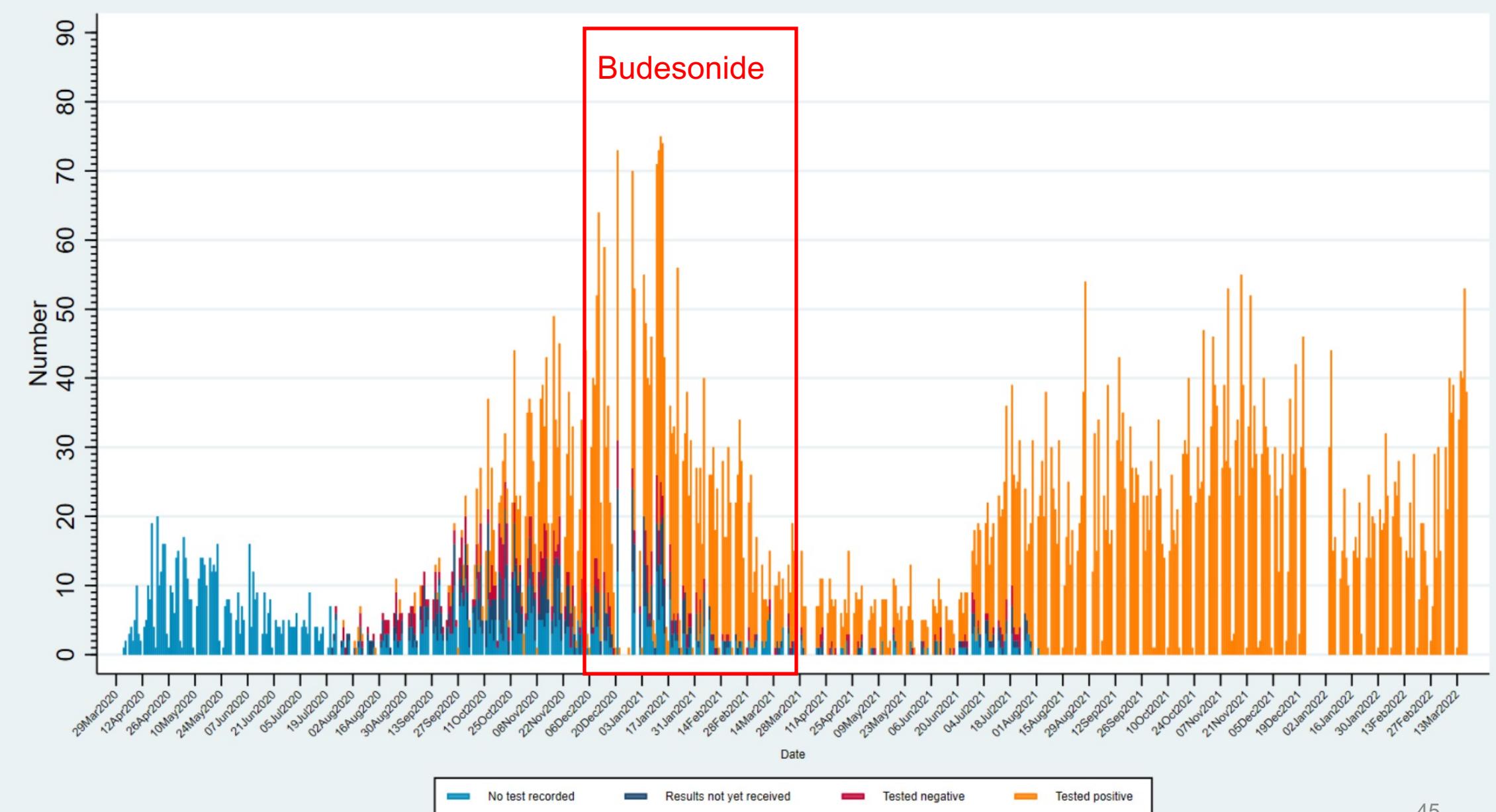
# Intervention Timeline in PRINCIPLE



# Intervention Timeline in PRINCIPLE

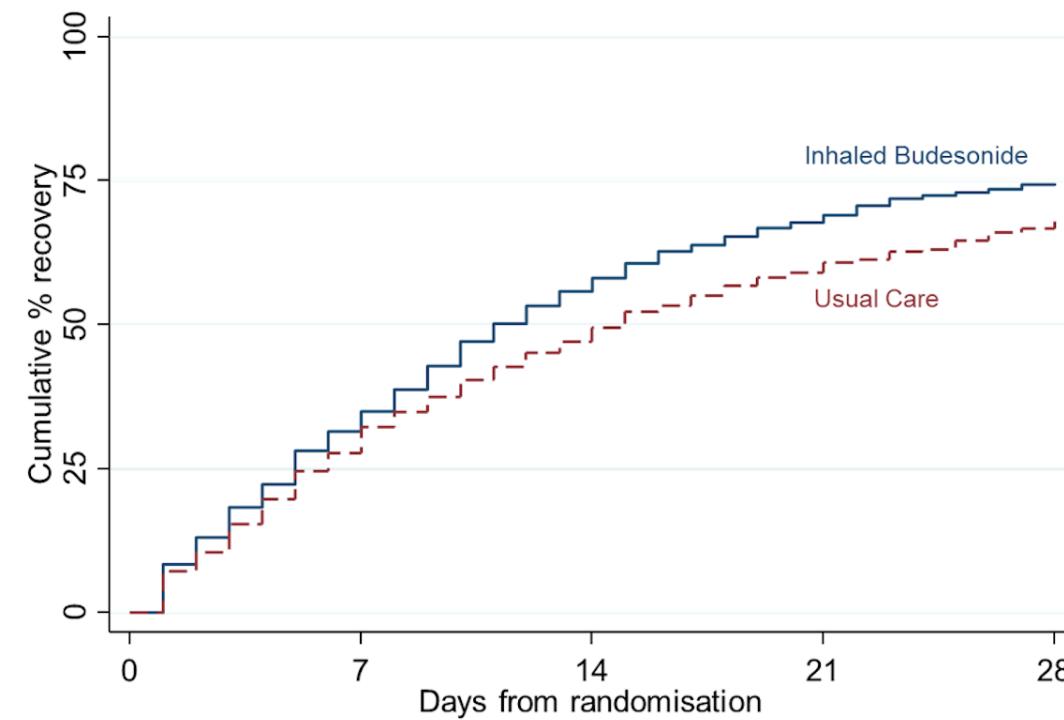


# Daily Randomization



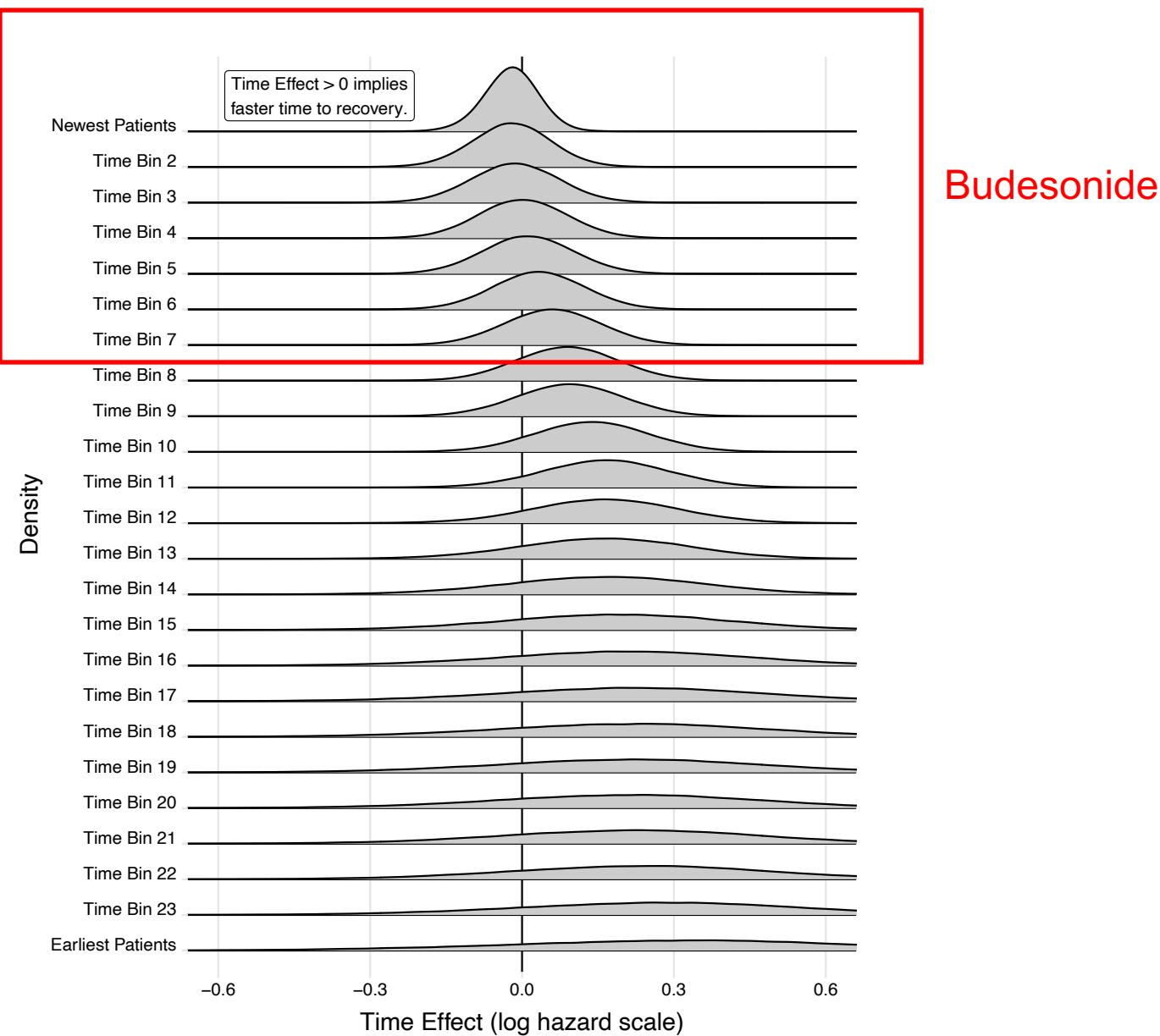
# Budesonide

- ~3 day benefit (Median TTR 12 vs. 15 days)
- ~2% reduction in hospitalization rate (7% vs. 9%)

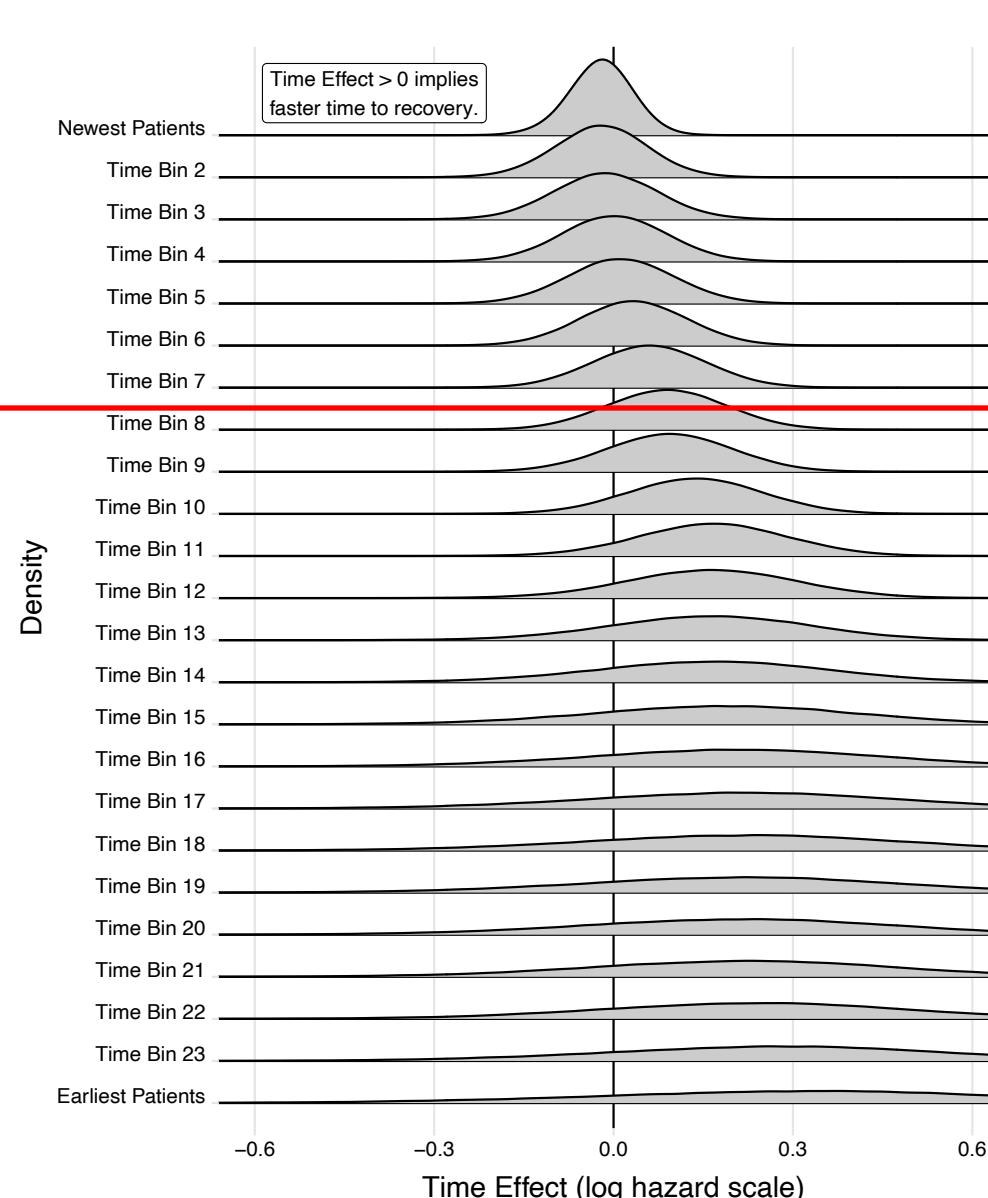


	Inhaled budesonide (95% BCI)	Usual care (95% BCI)	Estimated benefit median time to recovery or hospital admission or death rate (95% BCI)	Hazard ratio or odds ratio (95% BCI)	Probability of superiority
<b>Primary analysis—SARS-CoV-2-positive participants</b>					
Number of participants	787	1069	..	..	..
Time to first reported recovery, days*	11.8 (10.0 to 14.1)	14.7 (12.3 to 18.0)	2.94 (1.19 to 5.11)	1.21 (1.08 to 1.36)	>0.999
Hospital admission or death at 28 days†	6.8% (4.1 to 10.2)	8.8% (5.5 to 12.7)	2.0% (-0.2 to 4.5)	0.75 (0.55 to 1.03)	0.963
<b>Sensitivity analysis—concurrent randomisation population</b>					
Number of participants	787	838	..	..	..
Time to first reported recovery, days*	11.7 (9.8 to 14.2)	15.0 (12.5 to 18.3)	3.26 (1.46 to 5.43)	1.24 (1.10 to 1.39)	>0.999
Hospital admission or death at 28 days†	6.6% (3.8 to 10.1)	8.9% (5.2 to 13.1)	2.2% (0.0 to 4.9)	0.73 (0.53 to 1.00)	0.975

# Bayesian Time Machine



# Bayesian Time Machine



Budesonide

- Pre-specified Time Machine provided opportunity for additional precision of treatment effect
- Budesonide: Too much uncertainty of temporal drift and non-concurrent controls to benefit (small N, overlap, etc.)

# Bayesian Time Machine

- Incorporates ALL available data
  - Non-concurrent controls and other treatment arms
  - Generally better estimates, precision, and power
- Real-world applications
  - I-SPY2, GBM AGILE, Precision Promise, Healey ALS, REMAP-CAP, PRINCIPLE, etc.
- Platform trials are novel and complex!
  - Why do we insist on simple “unbiased” analyses?
    - Cost of lower precision and statistical power
  - Better estimation via modeling that leverages ALL platform trial data

**Modern analysis methods for modern trials!**