Platform trials – Only suitable for a pandemic?

Thomas Jaki

University of Cambridge
thomas.jaki@mrc-bsu.cam.ac.uk

University of Regensburg
thomas.jaki@ur.de

Collaborators

- Dominic Magirr, Novartis
- Peter Greenstreet, Ottawa Hospital Research Institute
- Tom Burnett, University of Bath
- Pavel Mozgunov, MRC BSU
- David Robertson, MRC BSU

Background

- A new disease brings many uncertainties
- Trials must start rapidly to identify treatments that
  - help patients
  - can be used as part of the outbreak response
- COVID-19 presentation is heterogeneous – can improve within days or last weeks or lead to death.

The 2014 Ebola experience

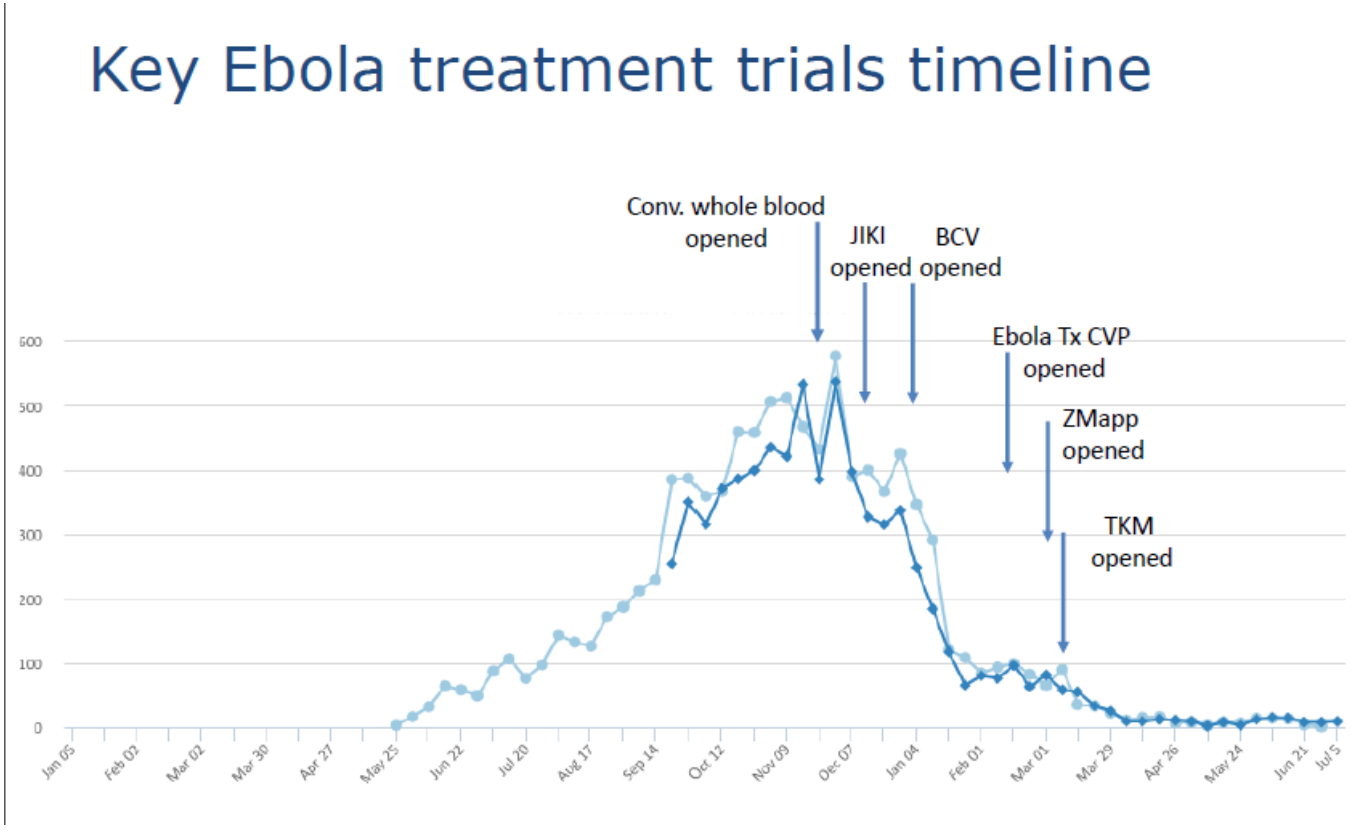


Figure: Number of new Ebola cases over time (courtesy of Sir Peter Horby).

LOTUS trial (Cao et al, 2020)

- Trial started on 18 January 2020
- Open-label trial of Loptinavir/Ritonavir
- Endpoint: rate of clinical improvement on 7-point scale up to day 28 or hospital discharge

7	Death
6	ICU, requiring ECMO and/or IMV
5	ICU/hospitalization, requiring NIV/HFNC therapy
4	hospitalization, requiring supplemental oxygen
3	hospitalization, not requiring supplemental oxygen
2	Not hospitalised, but unable to resume normal activities
1	Not hospitalised with resumption of normal activities

IMV, invasive mechanical ventilation; NIV, non-invasive mechanical ventilation.

- Trial started without formal design

Adapting the design

- Primary endpoint - informed by Influenza trials
  - changed to time to clinical improvement
- Formal sample size calculation
- Discussed group-sequential and multi-arm design options

Remdesivir trial (Wang et al, 2020)

- Group-sequential design
- 1 interim analysis planned at half-way point
- Primary endpoint: time to 2-point improvement on 6-point scale.
  - Collapsed category 1 and 2 from previous endpoint
- Recruitment within 10 days of symptom onset

The perfect trial during a pandemic

The perfect trial during a pandemic

- Uses meaningful and efficient endpoint
- Discards poor treatments quickly
- Can cope with multiple treatments
- Minimizes burden on frontline staff
- …

The perfect trial during a pandemic

- Uses meaningful and efficient endpoint
- Discards poor treatments quickly
- Can cope with multiple treatments
- Minimizes burden on frontline staff
- ...

$\Rightarrow$ is ADAPTIVE[1]

[1] Stallard et al (2020) for an overview of adaptive methodology

Idea of adaptive designs

Dimairo et al. (2020)

Modify an ongoing trial

by design

based on reviewing accrued data at interim

to enhance flexibility

Idea of adaptive designs

Dimairo et al. (2020)

Modify an ongoing trial

by design

based on reviewing accrued data at interim

to enhance flexibility



Idea of adaptive designs

Dimairo et al. (2020)

Modify an ongoing trial

by design

based on reviewing accrued data at interim

to enhance flexibility

without undermining the study's integrity and validity

Idea of adaptive designs

Dimairo et al. (2020)

Modify an on

by des

erim

without und

alidity



Adaptive Designs

Pallmann et al. (2018)

- Benefits: “better” and/or faster decisions
- Challenges:
  - Operationally more complex
  - Care needed around false decision risk
  - Risk of bias

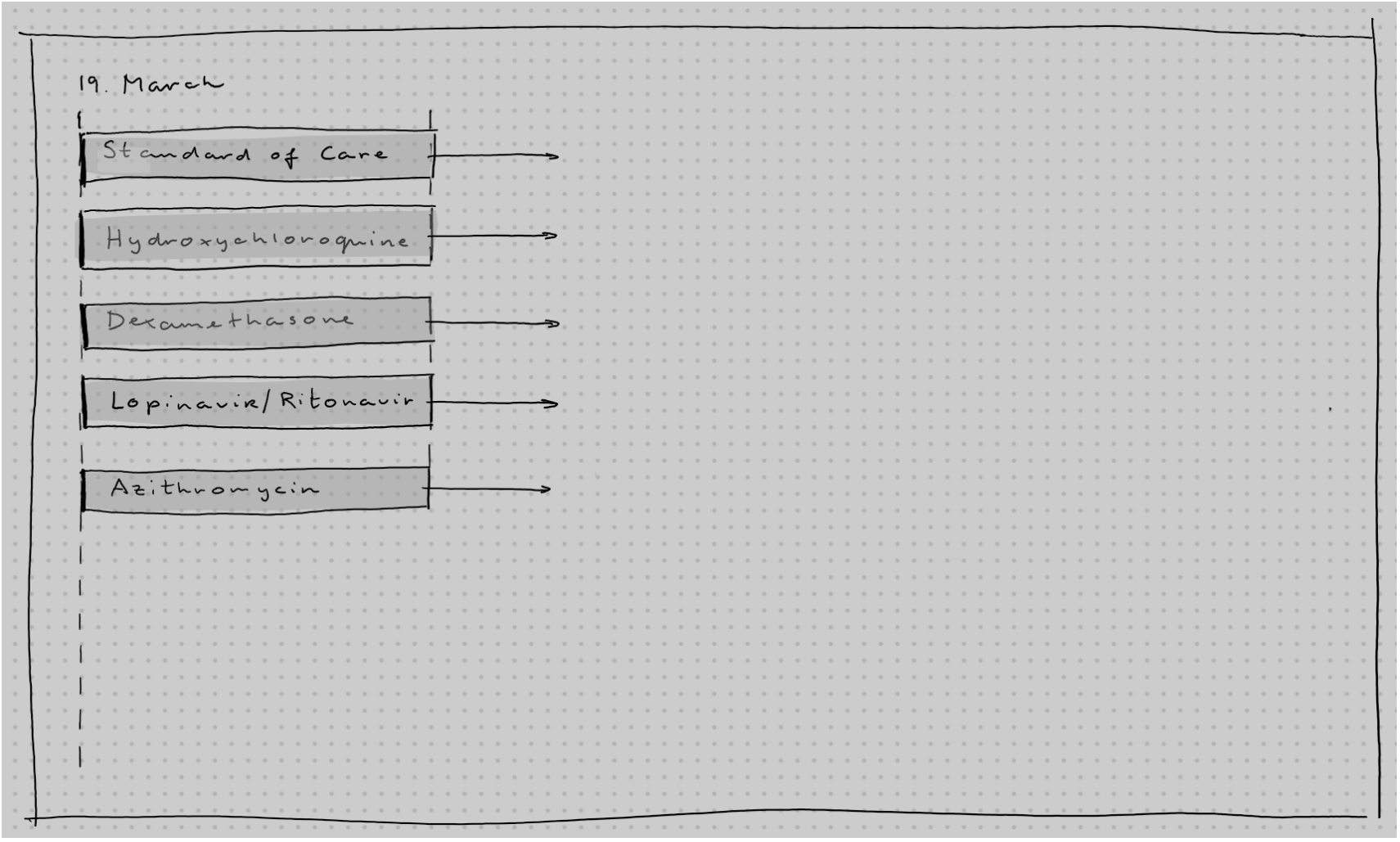
RECOVERY (www.recoverytrial.net)

- Multi-arm platform trial
  - Initially 4 treatments
- Endpoint: Mortality at 28 days
- Minimal data collection - data linkage

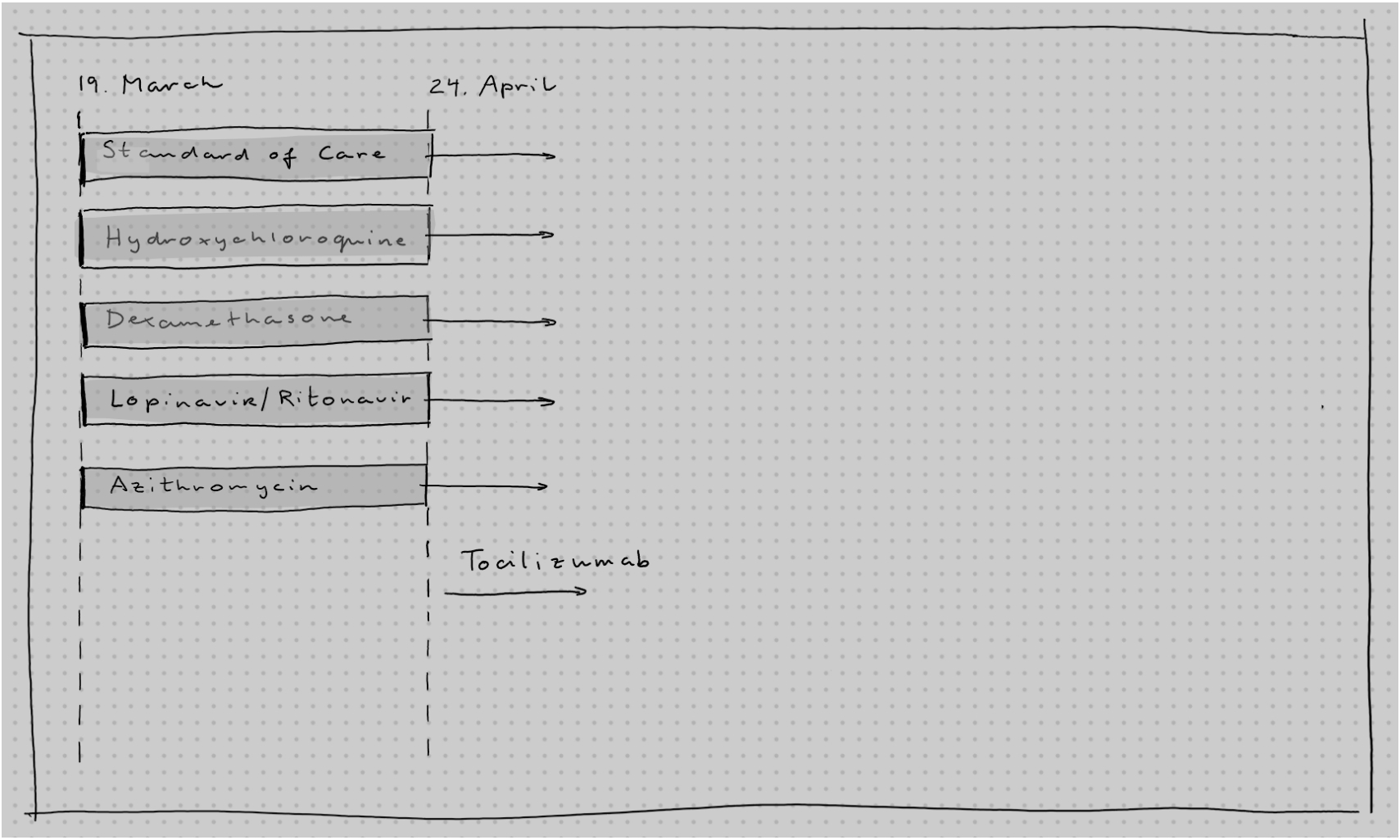
A great story . . .

- $\sim$ 11,500 patients recruited in 90 days
- $>$ 50,000 patients recruited to date
- 189 sites in 6 countries
- open to all ages
- 9 days from funding to first patient
- . . .

RECOVERY


19. March
Standard of Care
Hydroxychloroquine
Dexamethasone
Lopinavir/Ritonavir
Azithromycin

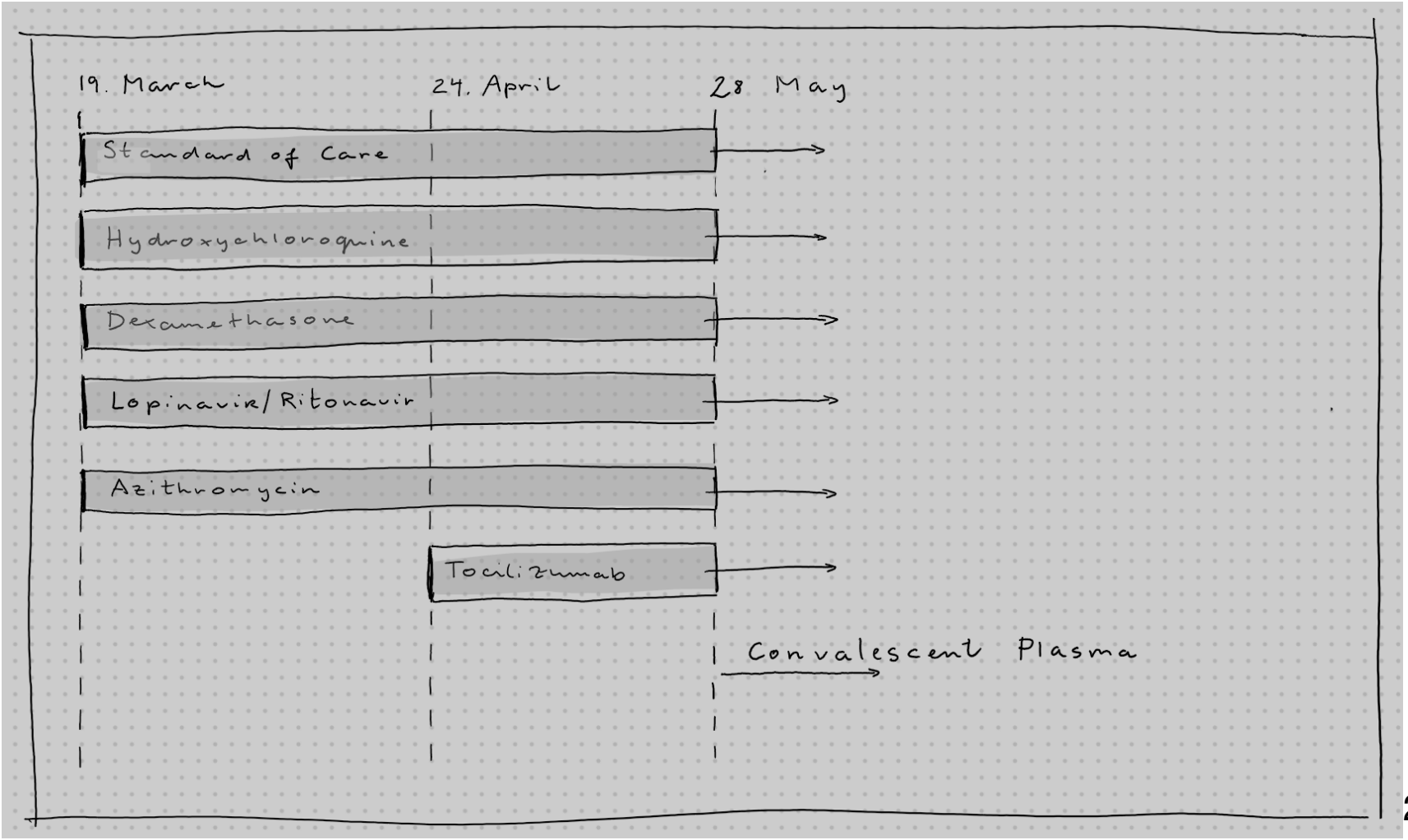
RECOVERY



[1]

[1] Horby et al. (2021a)

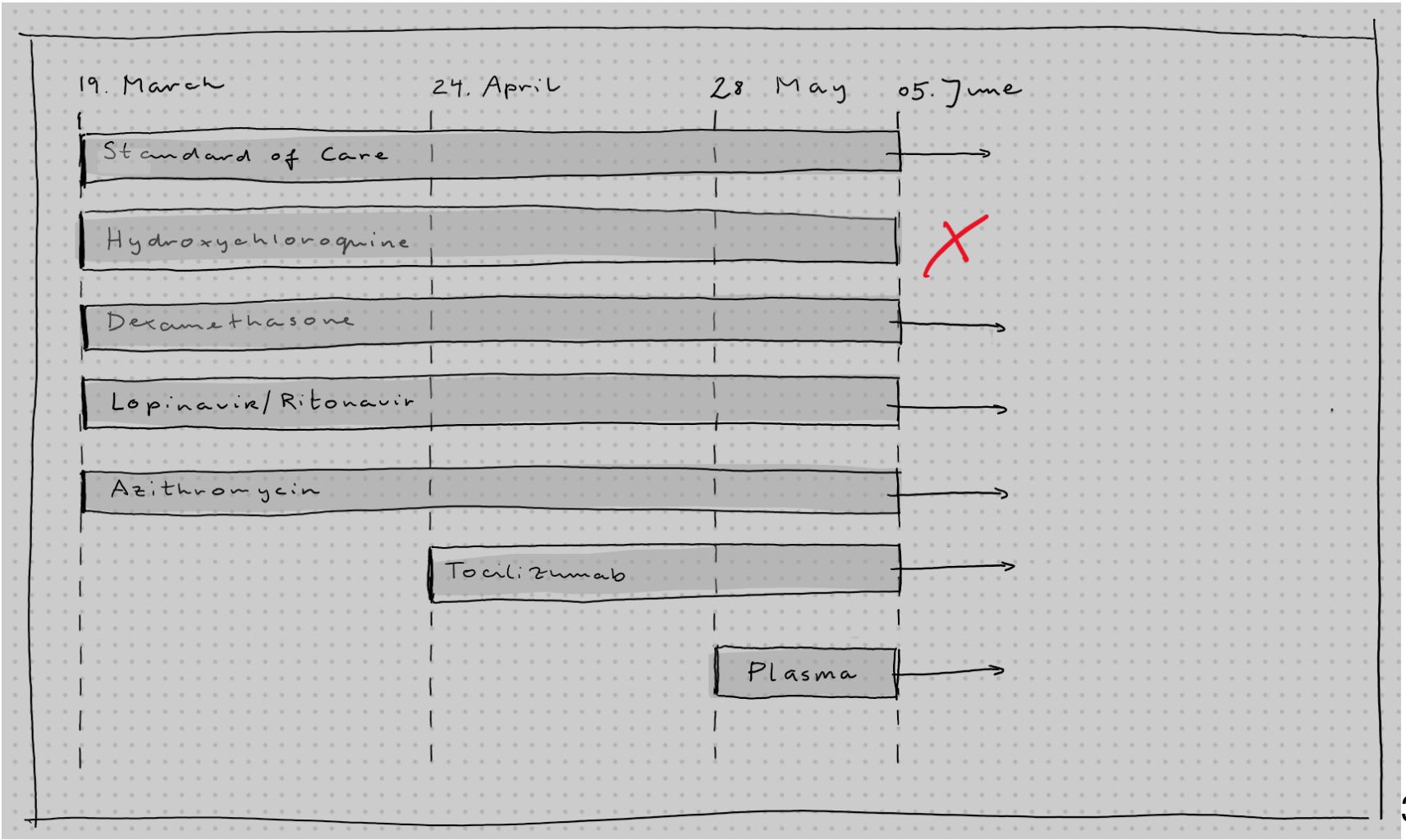
RECOVERY



[2]

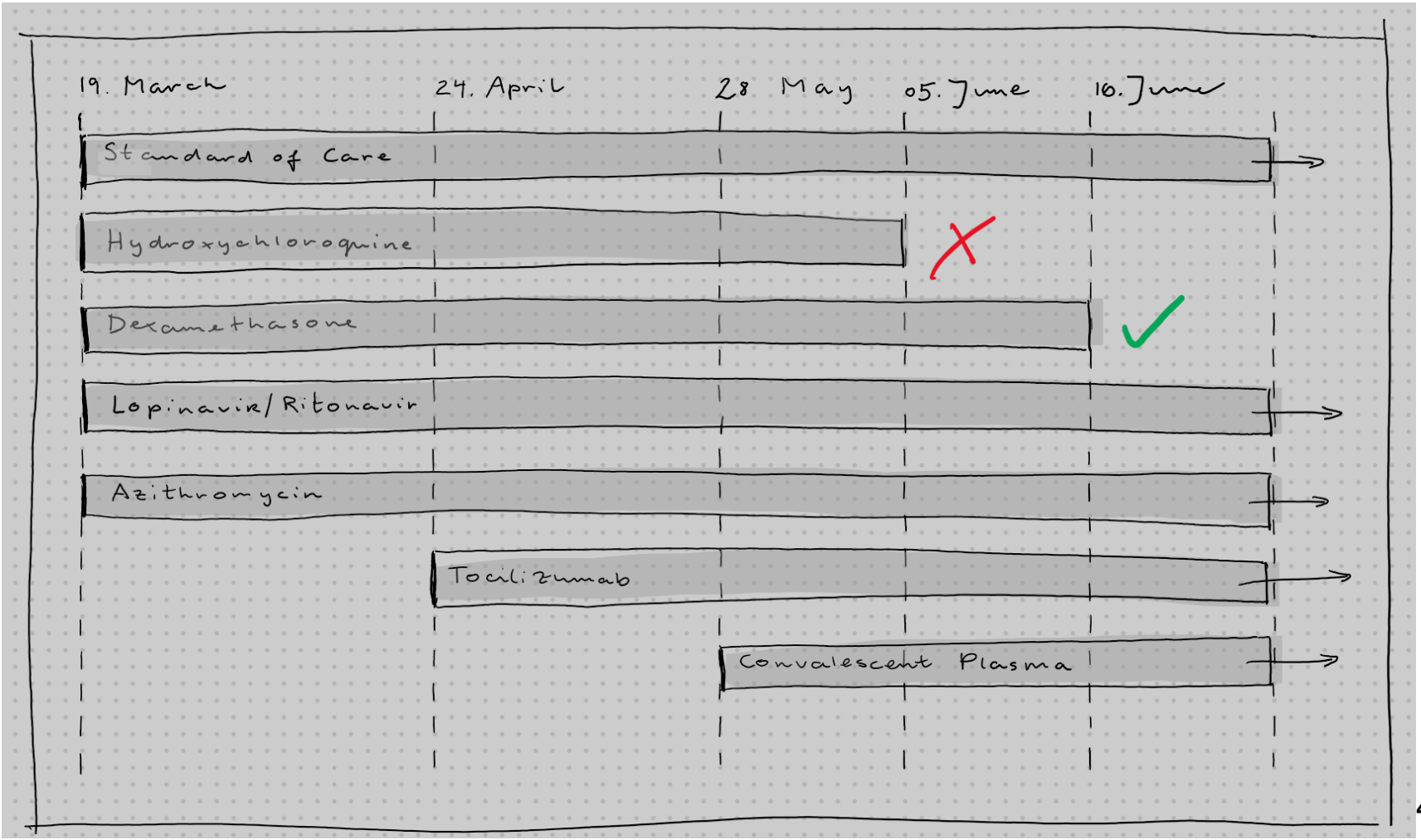
[2] Horby et al. (2021b)

RECOVERY



[3] Horby et al. (2021c)

RECOVERY

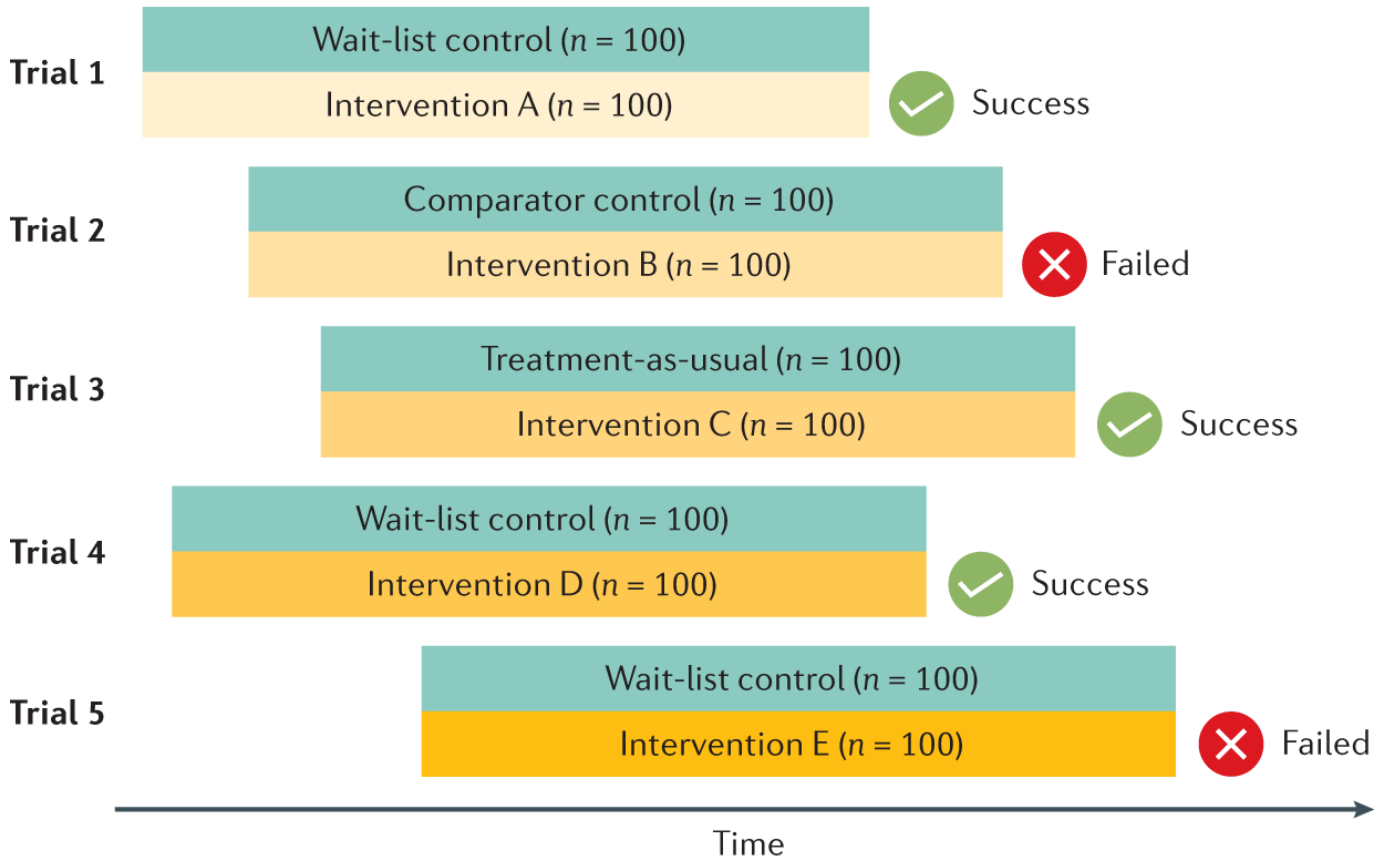


[4] Horby et al. (2021d)

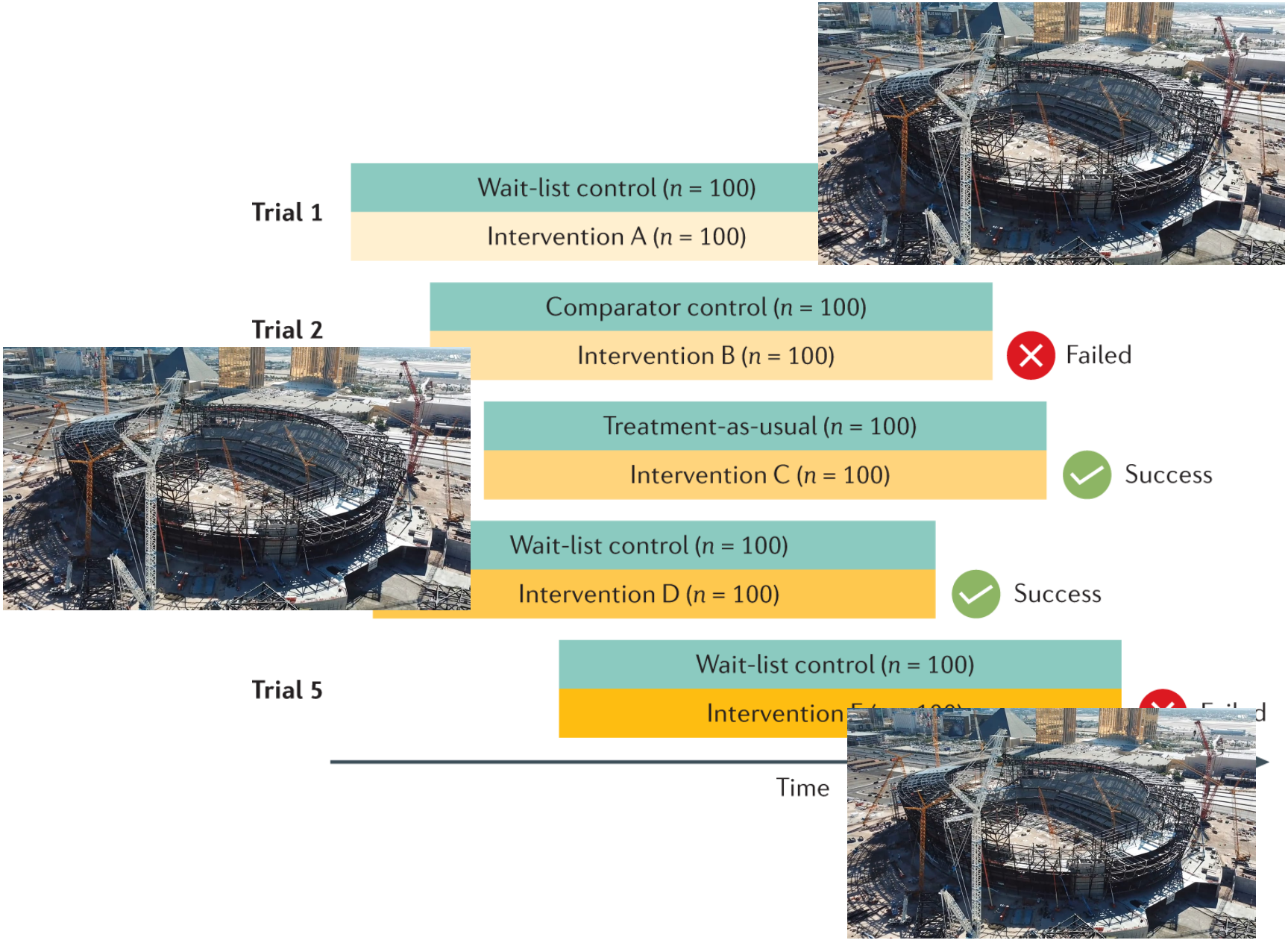
RECOVERY

- Answered 21 questions so far
  - 5 treatments reducing mortality
- Model/influenced several other (UK) trials
  - AGILE, PROTECT-CH, HEAL-COVID
- Trial now adapted to investigate flu and CAP
  - 1000 patients recruited from 15 countries in new indications

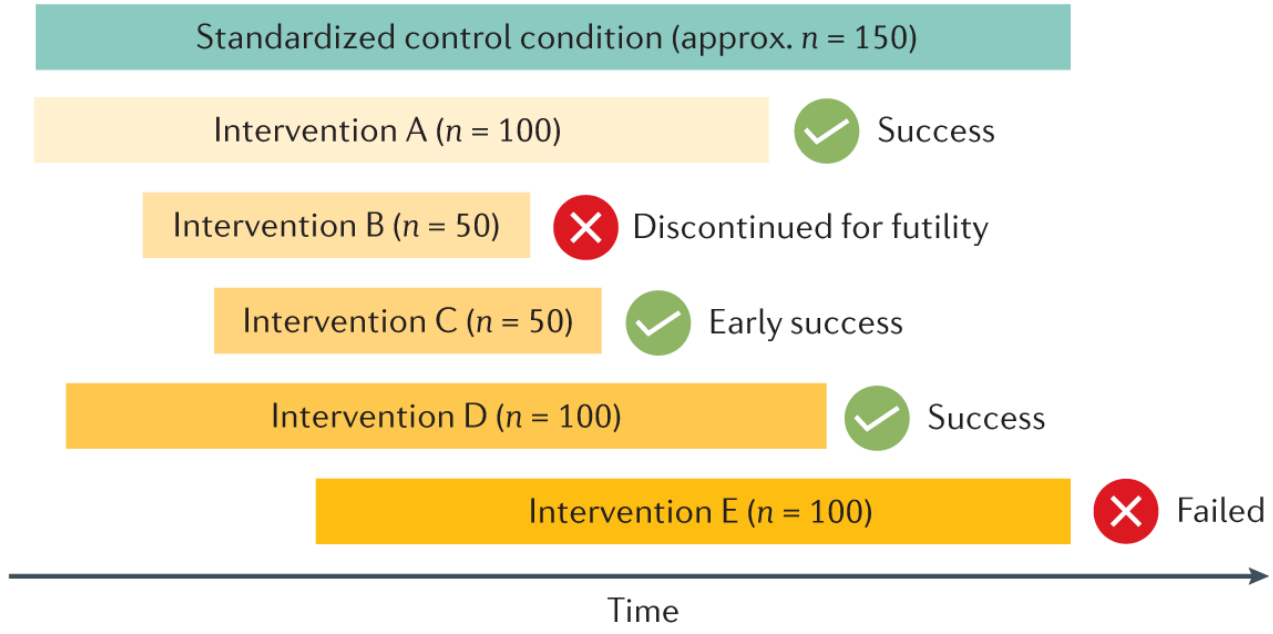
Idea of platform trials



Idea of platform trials



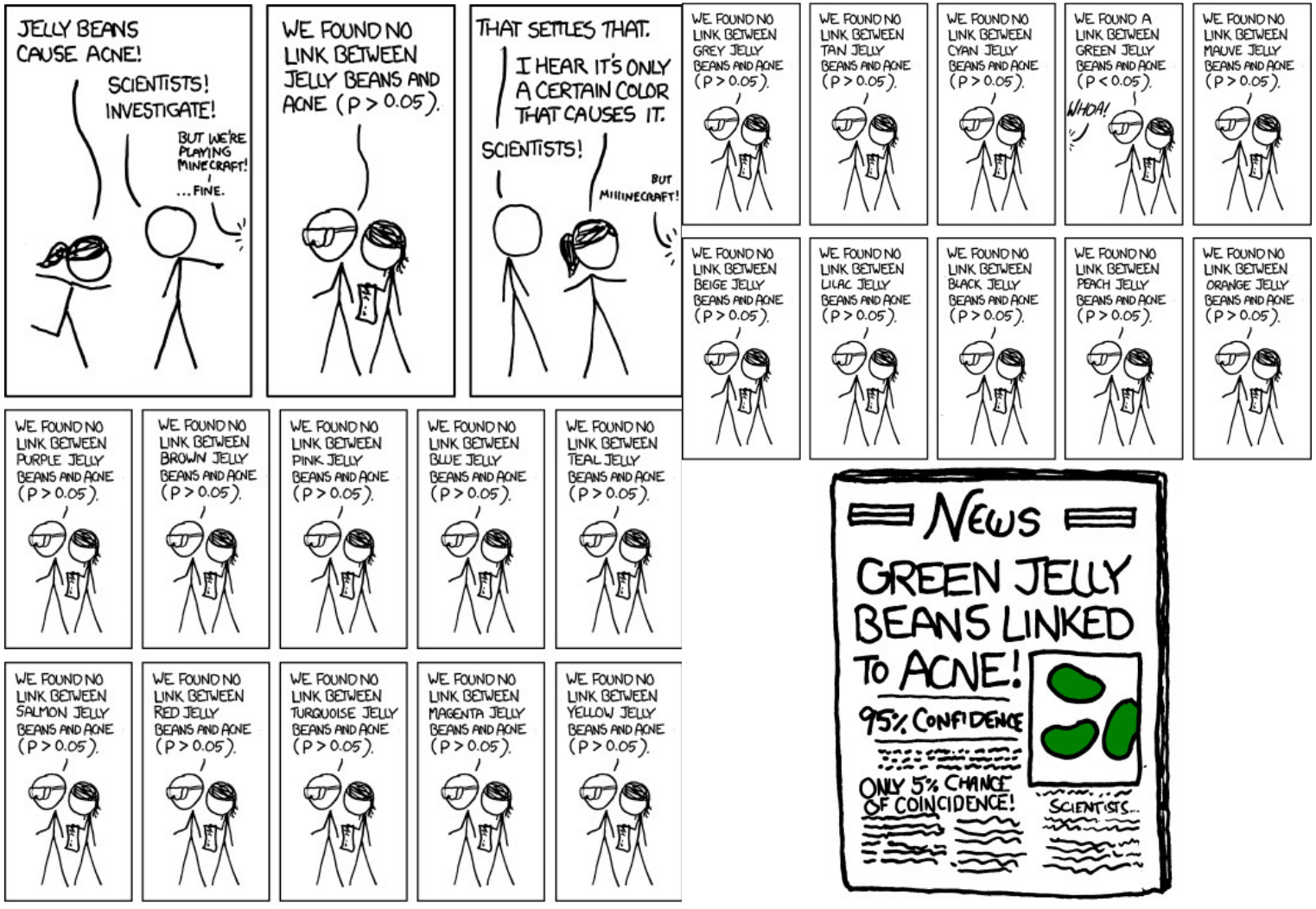
A platform trial



Platform trials

- Highly efficient
  - Use of infrastructure
  - Shared control group
  - Comparability of evaluations
  - Will stop investigations of bad treatments early
  - Enables moving forward with good treatments faster
- Popular especially during COVID-19 pandemic - but only suitable during a pandemic?

Multiple testing



https://xkcd.com/882/

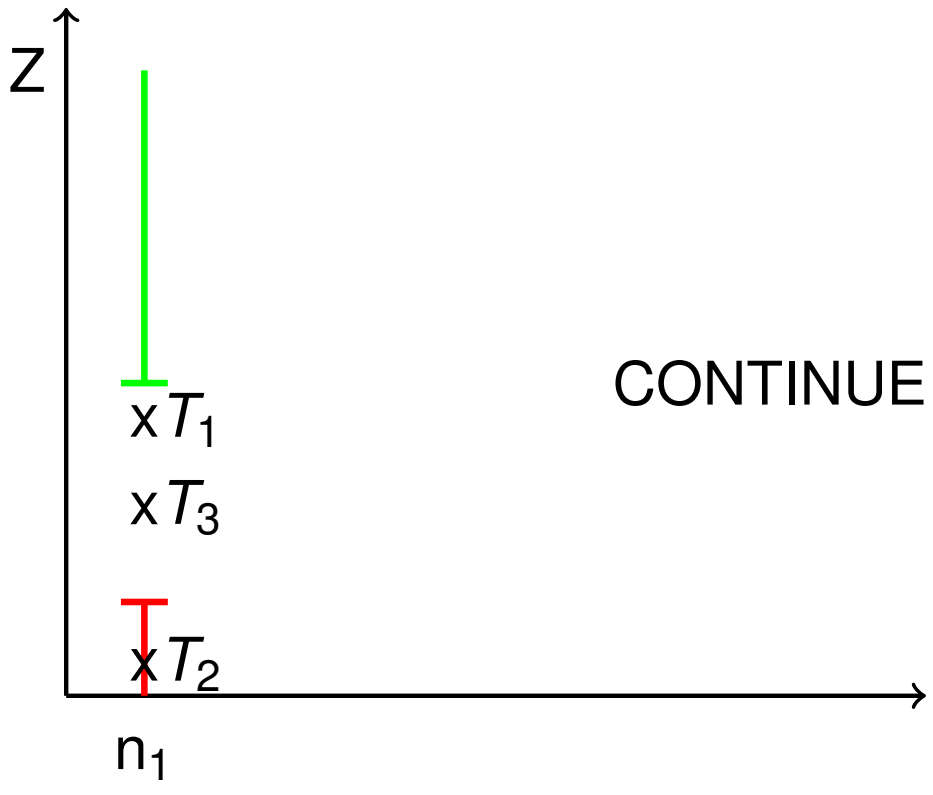
What you probably did not know

- Initial design of RECOVERY set to control FWER

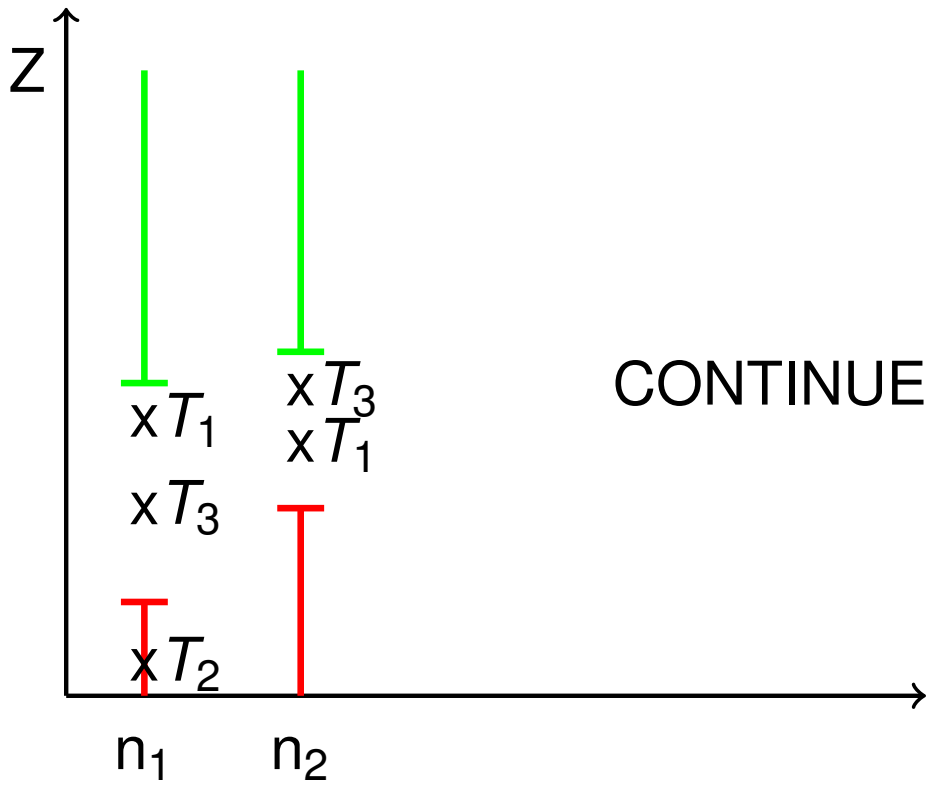
What you probably did not know

- Initial design of RECOVERY set to control FWER
- At the time not clear how to do so when adding arms

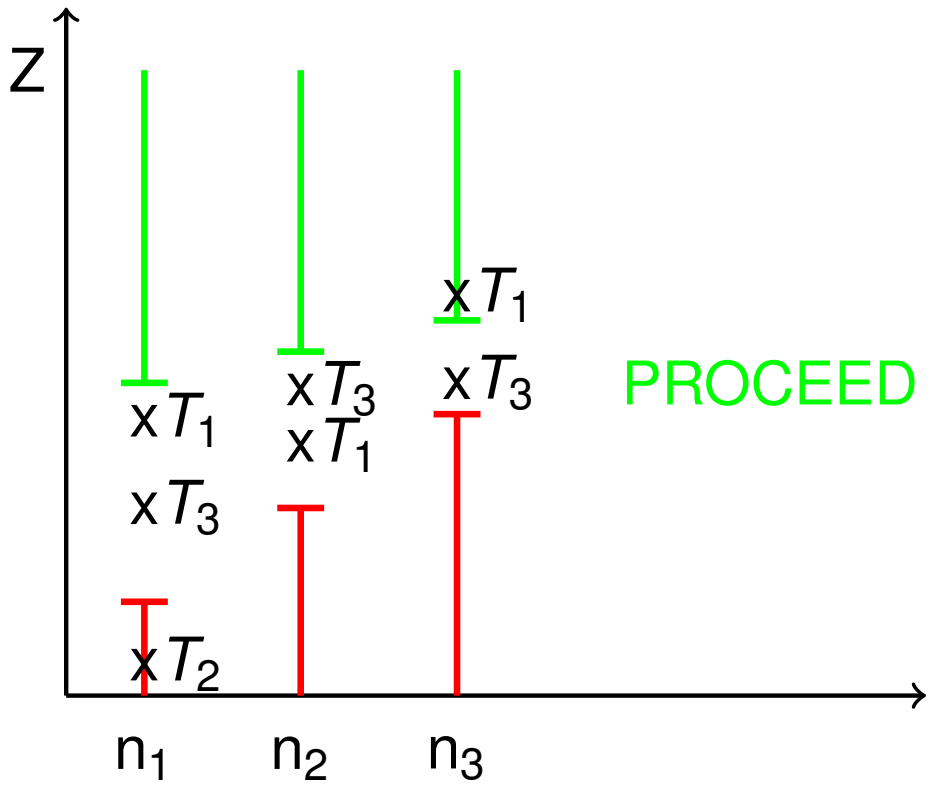
A multi-arm multi-stage (MAMS) design



MAMS



MAMS



Testing multiple hypothesis

Responses: $X_{k,i} \sim N(\mu_k, \sigma^2), \quad i = 1, \ldots, n, \; k = 0, 1, \ldots, K$

Individual null hypotheses:

$$\begin{aligned} H_1 : &\quad \mu_1 \leq \mu_0 \\ &\vdots \qquad\quad \vdots \\ H_K : &\quad \mu_K \leq \mu_0 \end{aligned}$$

Teststatistics: $Z_k = \frac{\bar{X}_k - \bar{X}_0}{\sigma\sqrt{\frac{2}{n}}}$ for $k = 1, \ldots, K$

Familywise error rate (FWER):

$$P(\text{reject at least one true } H_k) \leq \alpha$$

$\Rightarrow$ Can use Dunnett test (Dunnett, 1984)

More Multiple Testing

- J-stage trial $\Rightarrow$ up to KJ hypothesis tests.

Strong control of FWER

$$P(\text{reject at least one true } H_k) \leq \alpha$$

Weak control of FWER

$$P(\text{reject at least one true } H_k \mid H_G) \leq \alpha$$

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

Computing P(reject at least one true $H_k | H_G$)

(Magirr et al, 2012)

Problem: Test statistics are correlated due to the common control.

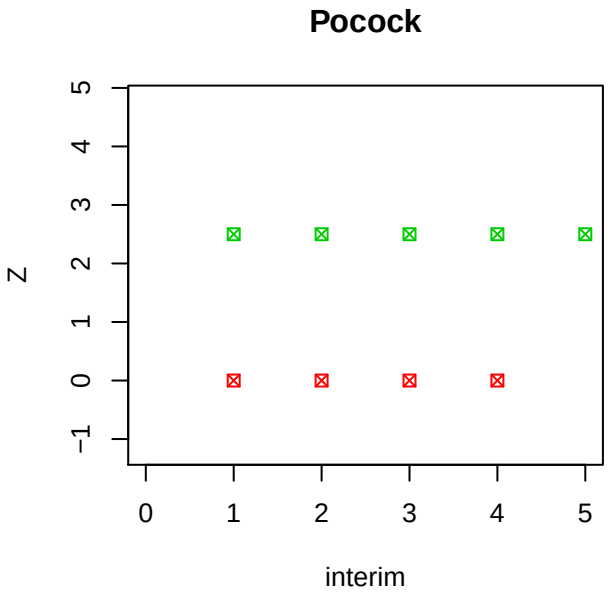
Solution: Condition on $\hat{\mu}_{0,J}$, the vector of sample means on control.

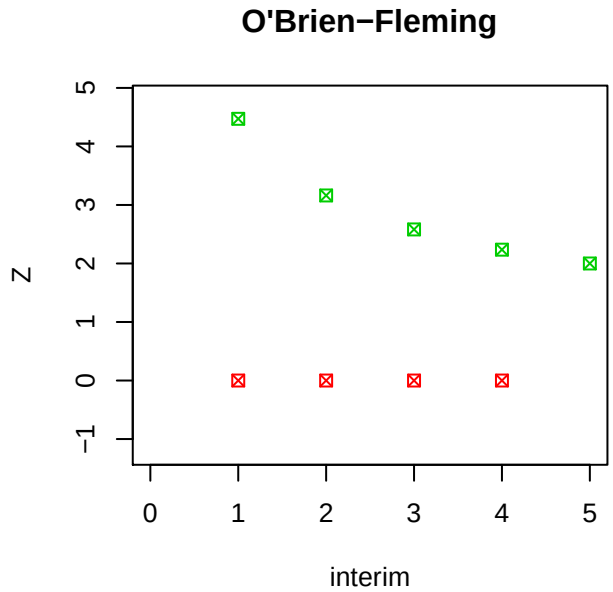
$$\alpha = 1 - \underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}_{J \text{ times}} \left[\sum_{j=1}^{J} P\left\{ \left(\bigcap_{i=1}^{j-1} B_{1,i} \right) \cap A_{1,j} \middle| \hat{\mu}_0,\ H_G \right\} \right]^K dF(\hat{\mu}_0)$$

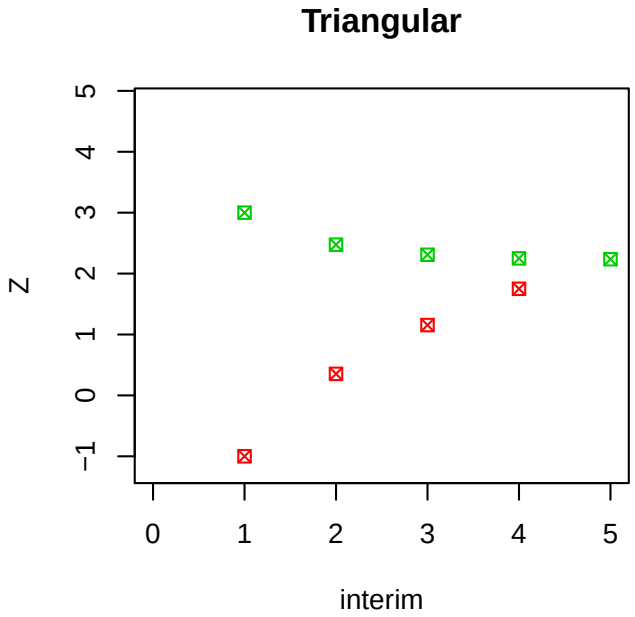
- $2J - 1$ unknowns $(l_1, \ldots, l_{J-1}, u_1, \ldots, u_J)$.

Boundary Constraints

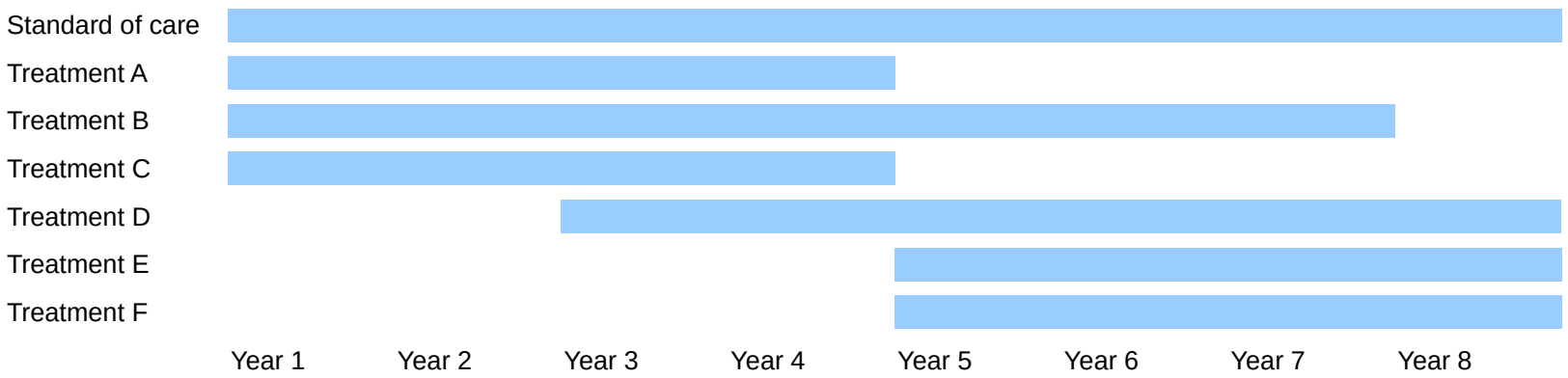
For $J > 1$ set $l_h = g(u_J)$ and $u_h = f(u_J),\ h = 1, \ldots, J-1$.







Multi-arm platform trials



Pre-planned adding (Greenstreet et al., 2024)

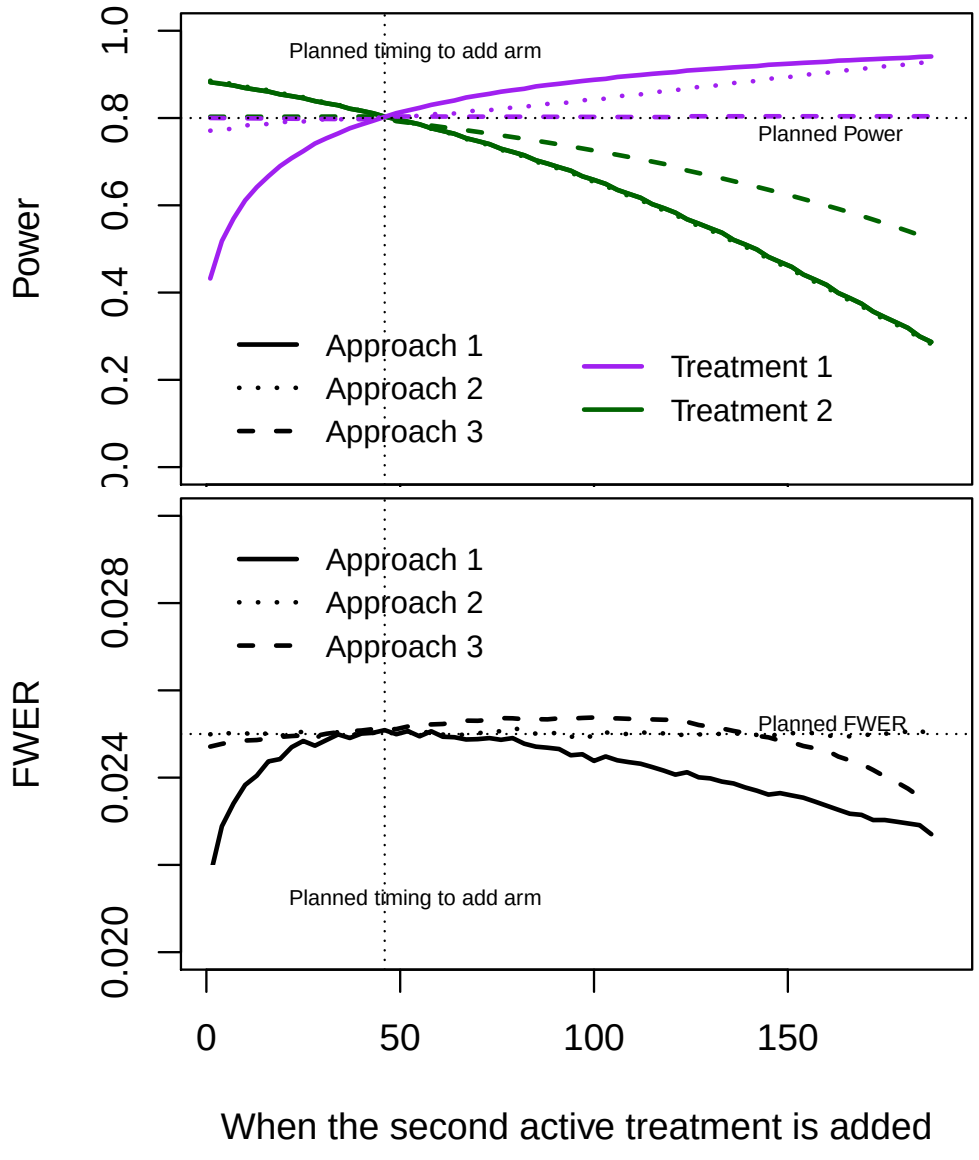
- Timing of adding an arm is known
- Same control group and endpoint
- Allocation ratio pre and post adding fixed (but possibly different)

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)

$\Rightarrow$ 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

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

Flexible adding (Burnett et al., 2025)

Initial design:

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

Modification:

- At first interim
- add 2 arms

Operating characteristics

	$\boldsymbol{\theta}$	$P(\boldsymbol{R})$	$\mathbb{E}_{\boldsymbol{\theta}}(N)$	$\mathbb{P}_{\boldsymbol{\theta}}(R_1)$	$\mathbb{P}_{\boldsymbol{\theta}}(R_2)$	$\mathbb{P}_{\boldsymbol{\theta}}(R_3)$	$\mathbb{P}_{\boldsymbol{\theta}}(R_4)$
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}, $\boldsymbol{R}$ is the event to reject at least one H_{0i}, $E_{\boldsymbol{\theta}}(N)$ is the expected sample size.

Some open questions

Adding arms

- Discussed methods for FWER control

Question

What are the right metrics to use in platform trials

Some open questions

Estimation

- Naïve ('maximum likelihood') estimators can over- or underestimate treatment effect
- Naïve confidence intervals could be too narrow or otherwise mislocated
- Recent ICH E20 guidelines on adaptive designs emphasizes the importance of appropriate post-trial estimation

Question

How to estimate effects in platform trials?

Some open questions

Changing comparator

- When changing the control group in a platform trial, it can be more powerful to discard previously collected information[5]

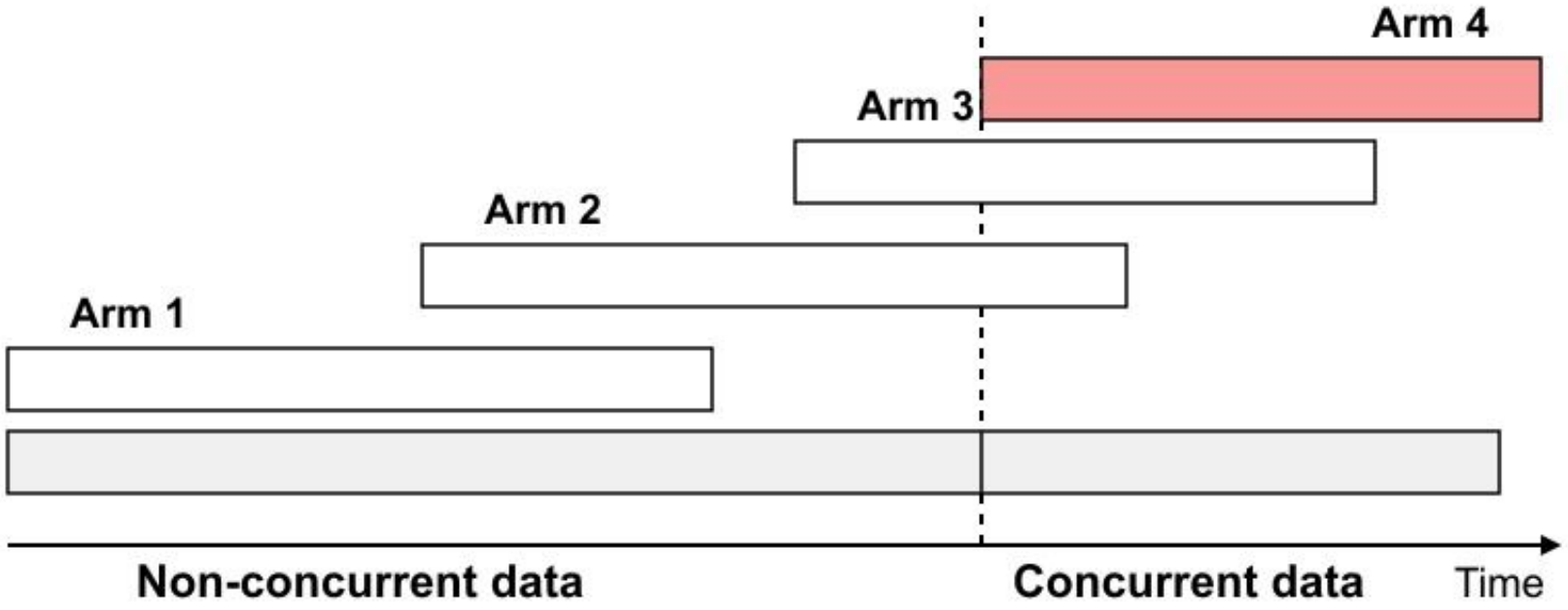
Question

How to efficiently use all data to make decisions

[5] Greenstreet P, **Jaki T**, ..., Mozgunov P (2025) Design of platform trials with a change in the control treatment arm. *Biometrics*. 81(2):ujaf073.

Other open questions

- Control group to use in analysis
- Control “treatment” to use in study
- Release of information
- Borrowing between arms
- …



Conclusions

- Platform trials were very popular during COVID-19 pandemic, but they can be very valuable outside of a pandemic
- Statistical method exist to deal with many of the additional complexities
- Platform trials are complex beasts and hence the following key questions need answering before embarking in them:
  - Will I actually add arms?
  - What is my control “treatment”

References

Bretz F, Koenig F. Commentary on Parker and Weir. (2020) *Clinical Trials*. 17(5):567-569.

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):3447-62.

Cao B et al (2020) A trial of lopinavir-ritonavir in adults hospitalized with severe Covid-19. *New England Journal of Medicine*.

Dimairo M et al. (2020) The Adaptive designs CONSORT Extension (ACE) Statement: A checklist with explanation and elaboration guideline for reporting randomised trials that use an adaptive design. BMJ. 369

Dunnett CW (1984) Selection of the best treatment in comparison to a control with an application to a medical trial. In *Design of Experiments: Ranking and Selection*

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.

Horby et al. (2021a) Tocilizumab in patients admitted to hospital with COVID-19 (RECOVERY): a randomised, controlled, open-label, platform trial. *The Lancet*. 397(10285):1637-1645.

Horby et al. (2021b) Convalescent plasma in patients admitted to hospital with COVID-19 (RECOVERY): a randomised controlled, open-label, platform trial. *The Lancet*. 397(10289):2049-2059.

Horby et al. (2021c) Effect of hydroxychloroquine in hospitalized patients with COVID-19. *New England Journal of Medicine*. 383:2030-2040.

Horby et al. (2021d) Effect of Dexamethasone in Hospitalized Patients with COVID-19 *New England Journal of Medicine*. 384:693-704.

König F et al (2008) Adaptive Dunnett Tests for Treatment Selection. *Statistics in Medicine* **27**:1612-1625.

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

Pallmann P et al. (2018) Adaptive designs in clinical trials: why use them, and how to run and report them. BMC Medicine. 16:29

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.

Stallard N et al. (2020) Efficient adaptive designs for clinical trials of interventions for COVID-19. *Statistics in Biopharmaceutical Research.*

Wang Y et al (2020) Remdesivir in adults with severe COVID-19: a randomized, double-blind, placebo-controlled, multicentre trial. *The Lancet*.