Outcome adaptive intervention roll out in stepped-wedge cluster randomised trials

Michael Grayling and Sofía Villar

SW-CRTs

- Routinely, talks on SW-CRTs would end with comments like:

  "Isn't it unethical to give the intervention to everyone when you don't know if it works?"

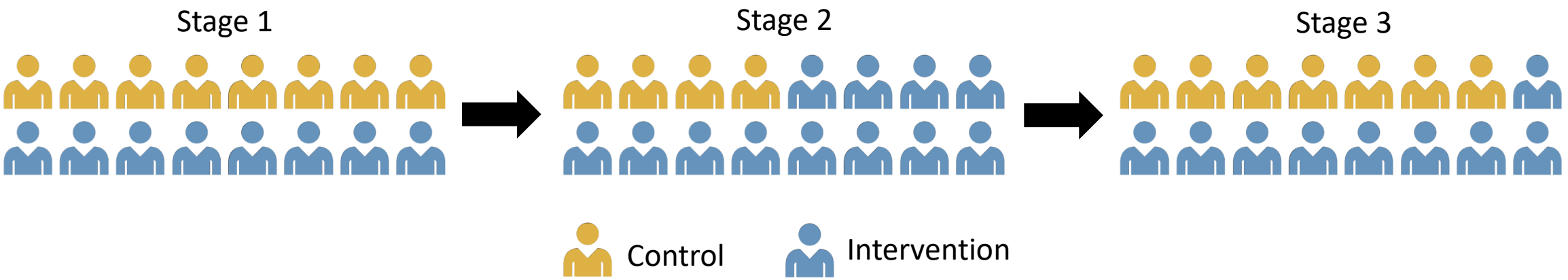
  "It's great everyone will get the intervention if it works, but what if it doesn't?"

- To which there were fairly standard responses:
  - Every cluster will get the intervention, but not necessarily every individual
  - You could choose sequences such that some clusters don't receive the intervention
  - Typically, we're not referring to pharmaceutical interventions, so usually concerns around safety/futility might not apply

- Nonetheless: maybe there are instances where this should concern us? If this is the case, what can we do?

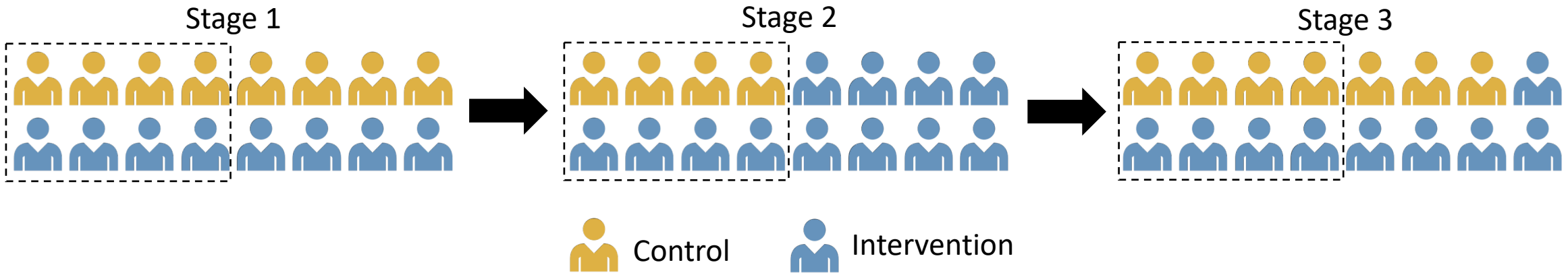
Outcome adaptive designs

- This is not dissimilar to a commonly discussed problem in early phase cancer trials
  - Historically non-randomised, with everyone receiving the experimental treatment
  - Now a large number of randomised trials with 1:1 allocation
  - More recently some work pushing towards using heavily skewed allocation ratios or…
- Outcome adaptive treatment allocation
  - Through a series of interim analyses, update the allocation ratio in favour of the treatment that is working best
  - Still monitor power, to permit reliable decision making



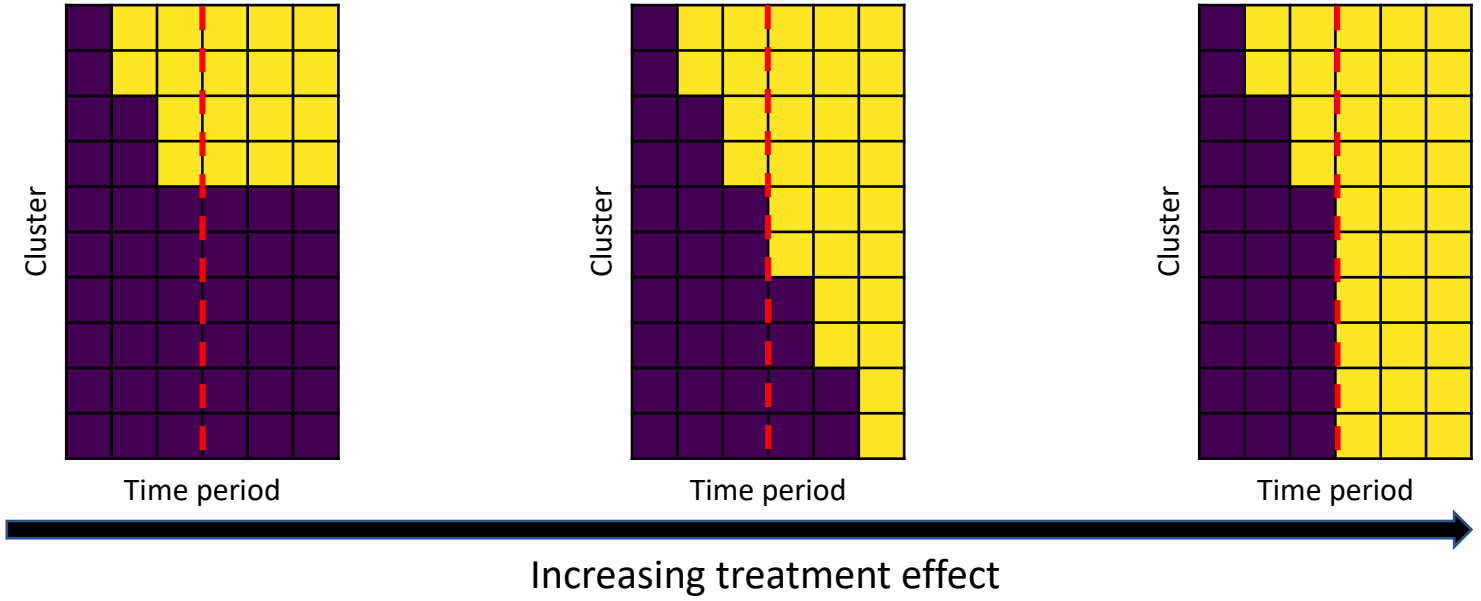
Outcome adaptive designs

- This is not dissimilar to a commonly discussed problem in early phase cancer trials
  - Historically non-randomised, with everyone receiving the experimental treatment
  - Now a large number of randomised trials with 1:1 allocation
  - More recently some work pushing towards using heavily skewed allocation ratios or…
- Outcome adaptive treatment allocation
  - Through a series of interim analyses, update the allocation ratio in favour of the treatment that is working best
  - Still monitor power, to permit reliable decision making



Outcome adaptive SW-CRTs

- So how should this work in a SW-CRT?

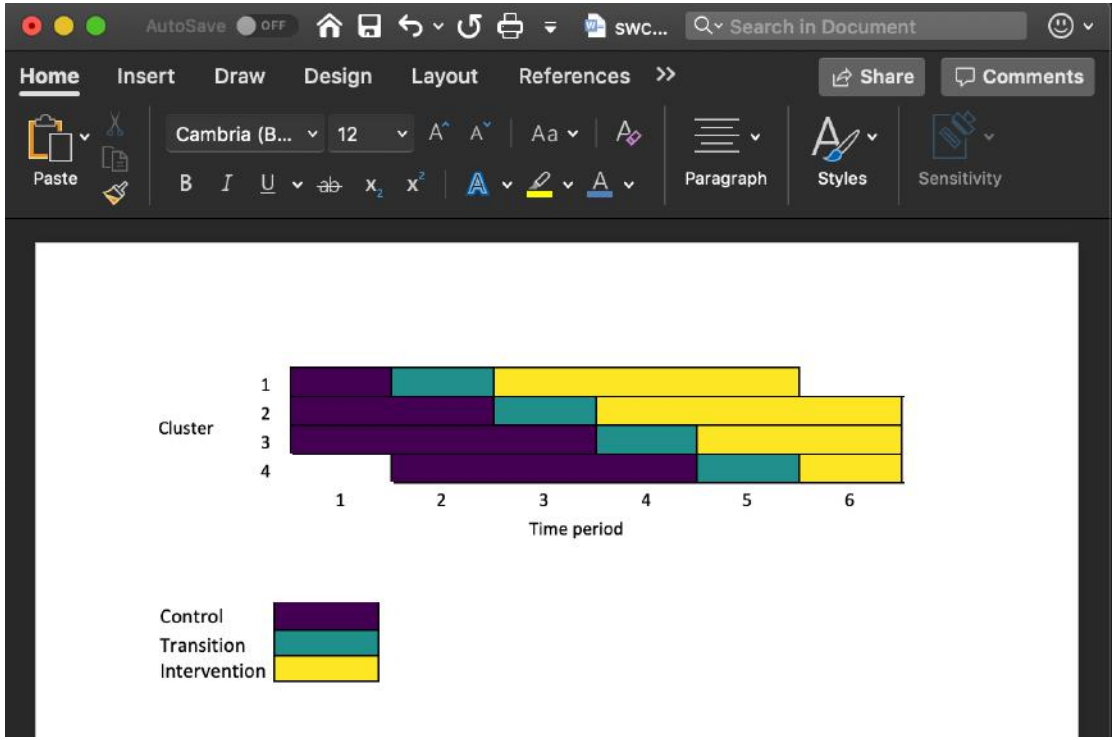


- Bigger question: how and will it work?
  - Treatment effect estimation more complicated: most of the literature for binary outcomes from individuals
  - We're making modifications for clusters already in the trial, as opposed to a new set of patients
  - A large number of interim analyses would be impossible
  - Can't undue 'mistakes' in modifications easily because of one directional switching
  - Potential lack of granularity to possible modifications

An aside: Drawing SW-CRT designs

- R function that draws SW-CRT designs given a simple matrix describing state in each cluster-period
- Can handle transition periods and incomplete block designs
- Outputs .pptx, .png, or .docx of the design

```
draw_swcrt.R
Source
1  sw ← rbind(c("C", "T", "I", "I", "I", "M"),
2             c("C", "C", "T", "I", "I", "I"),
3             c("C", "C", "C", "T", "I", "I"),
4             c("M", "C", "C", "C", "T", "I"))
5  draw_swcrt(sw)
6:1   (Top Level)   R Script
```



Hypothesis and analysis

- Specify set of time periods $\{t_1, t_2, \dots\}$ after which the roll out can be modified
- Specify treatment indicators X_{ij} for clusters $1, \dots, C$ for time periods $1, \dots, t_1$, and number of measurements m to gather per cluster-period
- Assume a linear mixed model will be used to analyse the data after time period $t = 1, \dots, T$:

$$\boldsymbol{y}_t = \boldsymbol{D}_t\boldsymbol{\beta}_t + \boldsymbol{Z}_t\boldsymbol{u}_t + \boldsymbol{\epsilon}_t$$

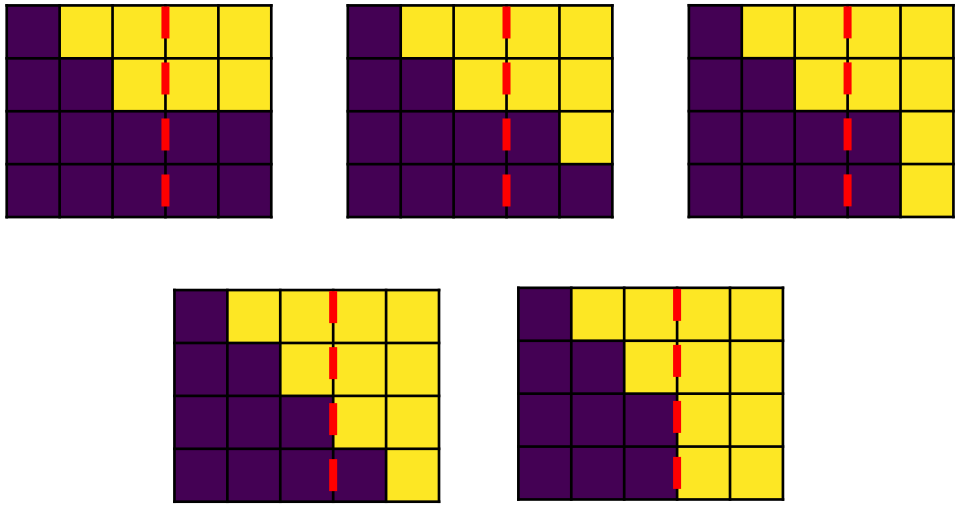
- Within $\boldsymbol{\beta}_t$ is a parameter for the intervention effect, θ. Use Wald test statistics:

$$Z_t = \frac{\hat{\theta}_t}{\sqrt{\mathrm{Var}(\hat{\theta}_t)}} = \hat{\theta}_t\sqrt{I_t}$$

- Level α one sided test at final analysis: $Z_T > \Phi^{-1}(1-\alpha)$. Desired power of $1-\beta$ when $\theta = \delta > 0$

Choosing a design: Balancing ethics and power

- At a particular interim analysis, we need a method for choosing between some list of possible choices for the X:
  - Can specify the possible choices for X in a number of ways, accounting for logistical or practical constraints
  - Require/not require everyone cluster to receive the intervention
  - Make sure the model is identifiable
  - Most efficient is placing no restriction on X



Choosing a design: Balancing ethics and power

- Need two functions. One to evaluate X based on ethical considerations, and one on power. Combine for an overall score:

$$SCORE(X) = wETH(X) + (1 - w)POW(X)$$

- Choose the X that maximises the score
- There is a logical option for the power function:

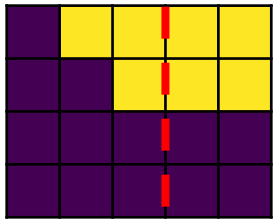
$$POW(X) = \frac{I_T(X)}{\max\limits_{\text{allowed } X'} I_T(X')}$$

where $I_T(X) = \left\{Var\left(\hat{\theta}_T | X\right)\right\}^{-1}$ is the information level at the final analysis

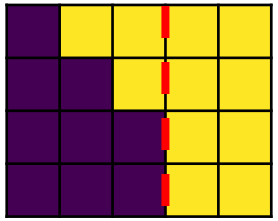
- Divisor here is because $ETH(\cdot)$ and $POW(\cdot)$ may not exist on the same scale

Ethical component: Desired features

- Some sensible features exist in generality
- It should recommend no additional clusters receive the intervention when $\theta \to -\infty$



- It should recommend every cluster receives the intervention immediately when $\theta \to \infty$



- In-between it should recommend some intermediate proportion of clusters receive the intervention
- Ideally, it should be monotonically increasing in the proportion of time spent in the intervention condition as θ increases

Ethical component: Simple option

- Specify a range of values for Z_t between which we use a balanced roll out:
  - $Z_t < l$ then stop the roll out
  - $Z_t > u$ then immediate roll out
  - $l \leq Z_t \leq u$ then use a balanced roll out
- Similar to including futility/efficacy stopping
- Does 'work' *but* to maintain power you typically can only terminate roll out/shift to immediate roll out at extreme effect sizes
- Need to recommend a broader range of designs, with the option for less certainty

Ethical component: Useful option

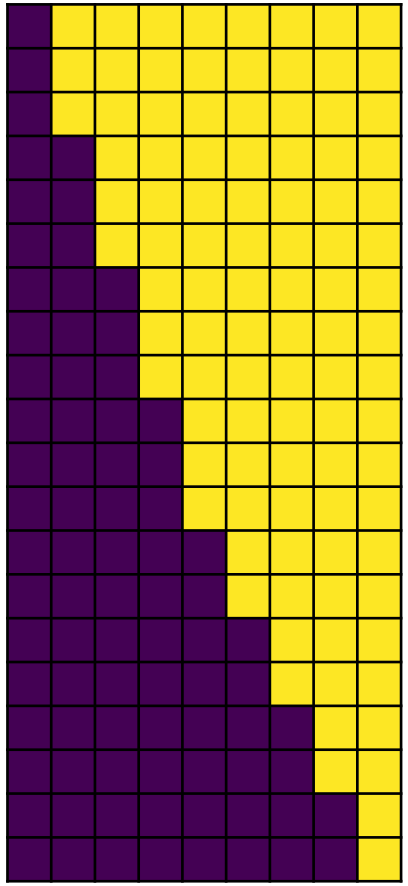
- Learn from previous approaches taken for block randomised designs
- Determine the proportion of clusters to allocate to the intervention in the next period using the density of a binomial distribution
- The number to switch, N, after time period t should follow the following distribution:

$$N \sim \text{Bin}\left[N_{\text{control}}, \Phi\left\{\frac{Z_t - m}{s\left(1 - \frac{t}{T}\right)}\right\}\right]$$

- Scale parameters m and $s > 0$ included:
  - m is the value of Z_t at which we recommend 50% switch
  - Increasing s allows us to skew recommendation towards 50%
- Score the ethics of X using the densities. E.g., if X switches one cluster $ETH(X) \propto \mathbb{P}(N = 1)$

Example: Design parameters

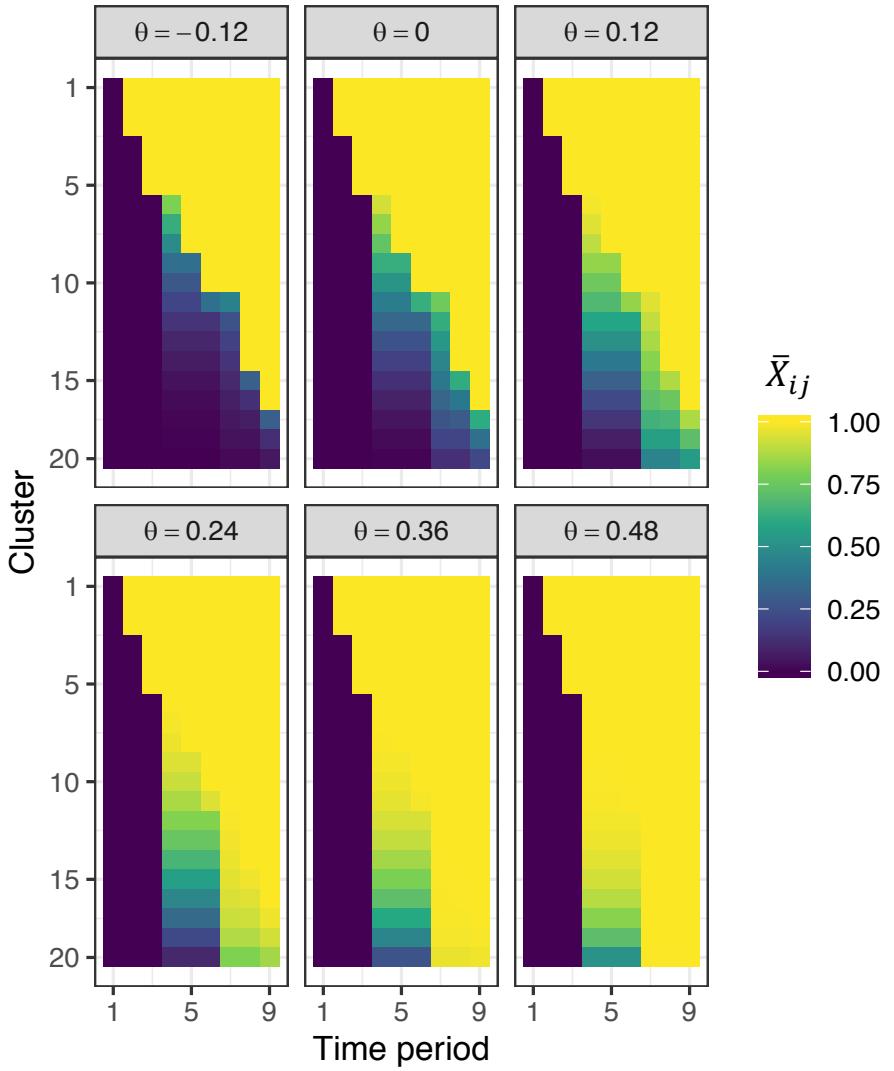
- Consider a cross-sectional trial analysed using the HH model:
  - $\rho = 0.1, \sigma^2 = 1.1$
- 20 clusters and 9 time periods. Fixed sample design has $\beta = 0.2$ for $\delta = 0.24$ with $\alpha = 0.05$
- Compare to outcome adaptive designs that can adapt following time periods 3 or 6
  - Using the same roll out for time periods 1-3
- All have $w = 0.5$, to give equal consideration to ethical/power components
- Rough rule: more analyses more efficient
  - Caveats to this if you analyse too early



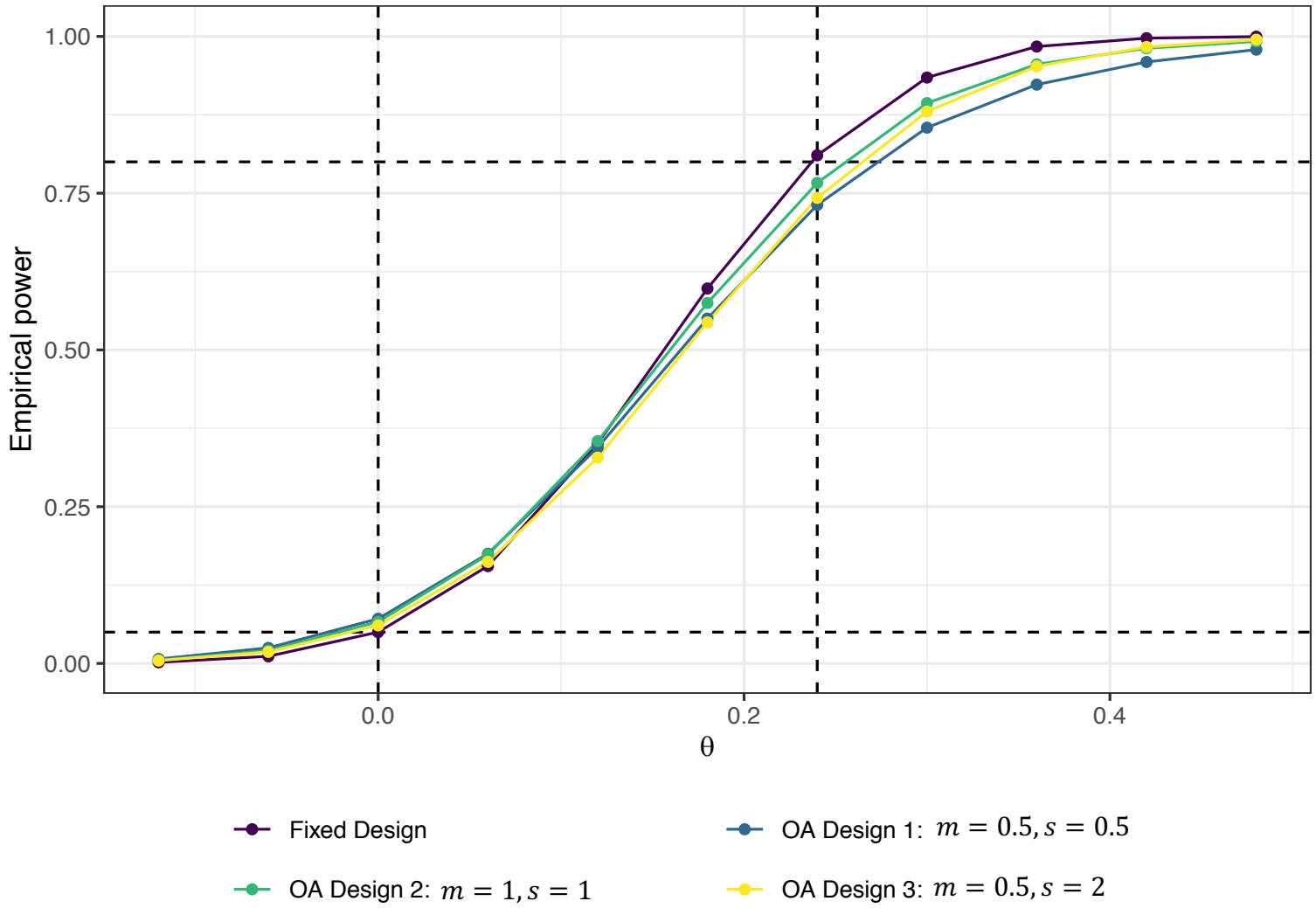
Example: Adaptive grid search

- To find a suitable design need to search over the parameters m and s
- Computationally intensive as simulation required
  - So actual algorithm to find best combination of m and s not feasible
- Hard to say in advance which combinations will lead to designs with desirable operating characteristics: adaptive grid search is helpful
- Looking to maintain type I and type II error-rates and decrease/increase roll out as appropriate
- Nonetheless, there is some redundancy in the search space
  - Give 3 examples

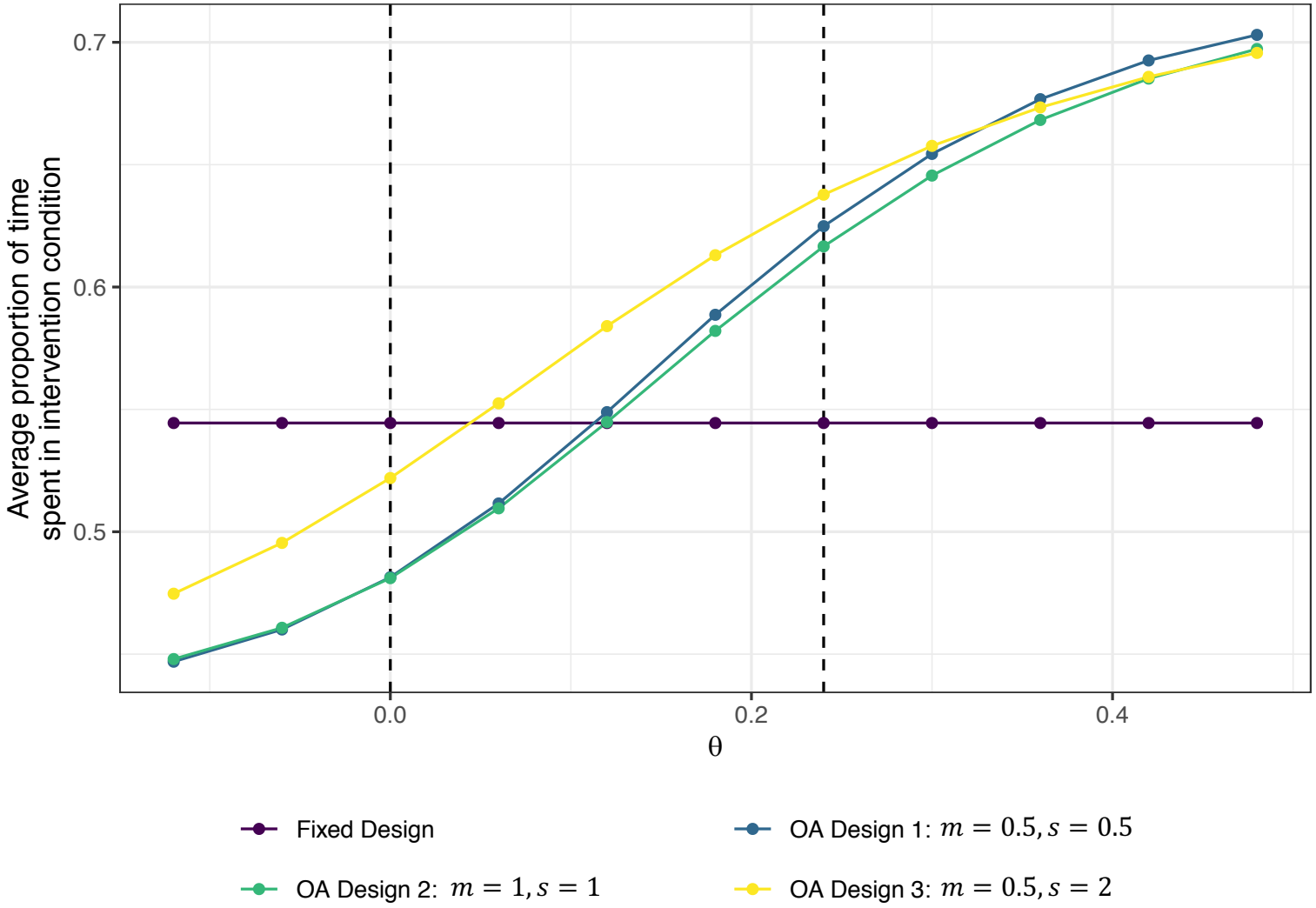
Average designs



Power


1.00
0.75
0.50
0.25
0.00
Empirical power
0.0
0.2
0.4
θ
Fixed Design
OA Design 1: $m = 0.5, s = 0.5$
OA Design 2: $m = 1, s = 1$
OA Design 3: $m = 0.5, s = 2$

Intervention allocation proportion



When might this be useful?

- Unlikely to be helpful/needed in many typical cases
  - Where the trial is an evaluation of a planned roll out
  - Where the intervention has limited safety concerns/costs associated with it
- Could be helpful if there are substantial differences in cost between the conditions
- Could also be relevant, for individually randomised stepped wedge, to certain rare diseases?
  - Don't 'spend' your potential pool of patients unless the intervention appears to work

When might this be useful?

- Like settings where outcome adaptive designs have been proposed to be useful, most likely relevant when there are substantial safety/ethical concerns
- SW-CRTs much discussed in context of evaluating vaccines during an epidemic:

Quantifying ethical tradeoffs for vaccine efficacy trials during severe epidemics

Steven E. Bellan[*1,2], Juliet R.C. Pulliam[3,4], Rieke van der Graaf[5], Spencer J. Fox[6], Jonathan Dushoff[7], Lauren Ancel Meyers[6,8]

Vaccine testing for emerging infections: the case for individual randomisation

Nir Eyal,[1] Marc Lipsitch[2]

Issues in the Use of Stepped Wedge Cluster and Alternative Designs in the Case of Pandemics

Ingeborg van der Tweel [a] & Rieke van der Graaf [a]

Novel Ordered Stepped-Wedge Cluster Trial Designs for Detecting Ebola Vaccine Efficacy Using a Spatially Structured Mathematical Model

Ibrahim Diakite[1]*, Eric Q. Mooring[2], Gustavo E. Velásquez[3,4,5], Megan B. Murray[1,2,6,7]*

Deciphering assumptions about stepped wedge designs: the case of Ebola vaccine research

Adélaïde Doussau, Christine Grady

Statistical power and validity of Ebola vaccine trials in Sierra Leone: a simulation study of trial design and analysis

Steven E Bellan, Juliet R C Pulliam, Carl A B Pearson, David Champredon, Spencer J Fox, Laura Skrip, Alison P Galvani, Manoj Gambhir, Ben A Lopman, Travis C Porco, Lauren Ancel Meyers, Jonathan Dushoff

Emergency Ebola response: a new approach to the rapid design and development of vaccines against emerging diseases

Claire M Tully, Teresa Lambe, Sarah C Gilbert, Adrian V S Hill

Stepped-wedge trial design to evaluate Ebola treatments

**Jolanta Piszczek, Eric Partlow*

When might this be useful?

- Actually a setting where the resources would be in place to carry out an interim analysis
- Outcome is also quickly evaluated, which is *very* important to adaptive design:

"a shorter time between intervention and outcomes is expected, such as in the case of an Ebola outbreak"

Piszczek and Partlow (2015)

- You do want to give the intervention as quickly as possible if it works, but perhaps not if it doesn't:

"At the time SW designs were of potential interest for studying experimental Ebola vaccines, there were **very few data about the safety** or immunogenicity of any of the vaccine candidates. Therefore, **proposals to use SW designs were based on hope** (rather than preliminary data) that the vaccines might do more good than harm."

"Furthermore, when considering the use of SW design for experimental drugs or vaccines, **interim analysis** and stopping rules for efficacy, safety and futility **might be necessary**"

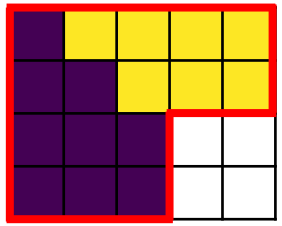
Doussau and Grady (2016)

"If little or nothing is known about this intervention, it **may turn out as equivalent to control**, or at least not superior. In case of inferiority of the intervention, clusters are exposed longer than perhaps necessary to the inferior treatment"

van der Tweel and van der Graaf (2013)

Summary

- It is possible to reliably adapt the roll out in an ethical manner, without compromising (substantially) on the type I and type II error-rates
  - Surprising!
  - Even true for small C and T
- Aided by the fact that there are often a wide range of designs that meet the power requirements
- There is also a natural way to control the allocation to the two arms



- Desirable performance also possible with a limited number of interim analyses
- Nonetheless, would be complicated in practice. More work needed to precisely assess utility