Adaptive design of cluster randomised trials

Michael J. Grayling

MRC Biostatistics Unit
Cambridge, UK

2018 HTMR Network Annual Meeting

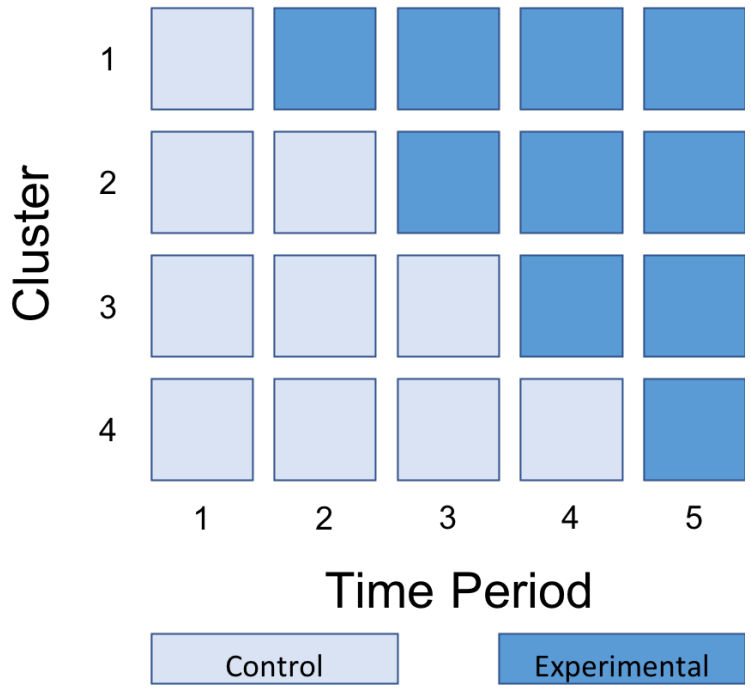
Outline

Adaptive design of PG-CRTs

- Much research has now been conducted to facilitate adaptive trial design
- But the majority of focus has been on individual level randomisation
- Some exceptions for parallel group (PG) cluster randomised trials (CRTs)
  - Lake *et al* (2002) discussed re-estimation of the required number of clusters
  - Zou *et al* (2005) described group sequential design for binary outcomes
  - van Schie and Moerbeek (2014) considered re-estimation of the required number of individuals per cluster

CRXO and SW-CRTs

- In recent years there has been increased interest in cluster randomised crossover (CRXO) and stepped-wedge (SW) CRTs



- Motivation comes from considerations related to these designs

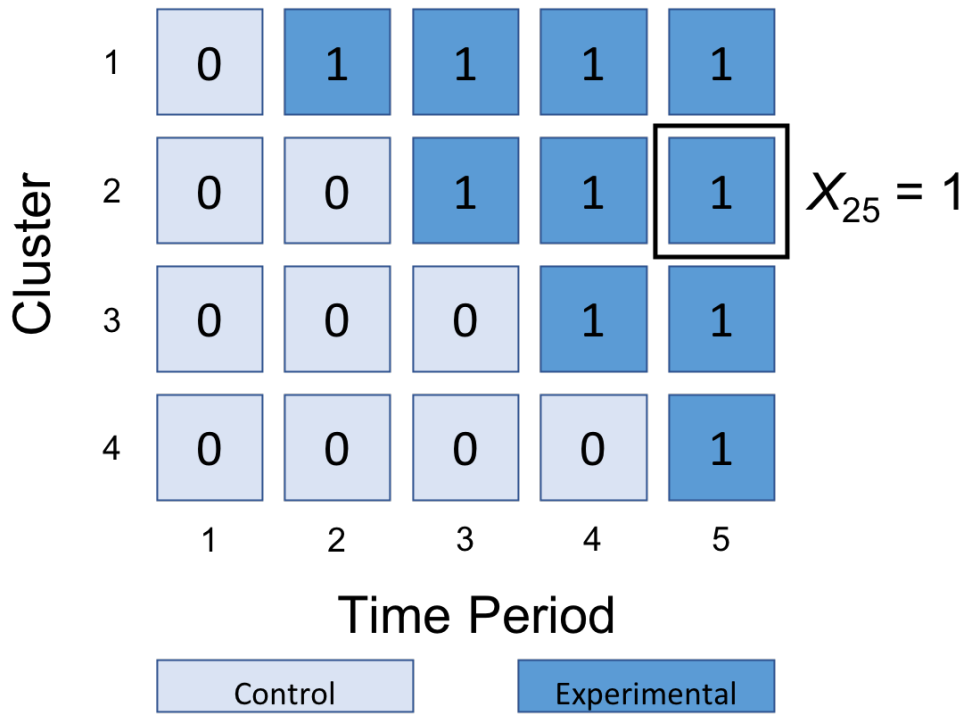
CRXO and SW-CRTs

- 31% of SW-CRTs failed to find efficacy on their primary outcome
- Possibly associated with over enthusiastic use of the design
- Or could be due to the challenge of specifying the variance components
- Describe a flexible framework for incorporating interim assessments of futility/efficacy in SW-CRTs/CRXO trials
- Detail how we can re-estimate the required sample size in a blinded or unblinded manner

Outline

Trial setting

- Consider a SW-CRT in C clusters over T time periods, with m measurements per cluster per period
- Suppose the treatment allocations have also been specified, via a matrix X of zeroes and ones



Analysis

- Can be used when there are different participants in each period, or the same participants
- They also work for almost any choice of linear mixed model
- For simplicity

$$Y_{ijk} = \mu + \pi_j + \tau X_{ij} + c_i + e_{ijk}.$$

- The random cluster effect $c_i \sim N(0, \sigma_c^2)$, and the residual error $e_{ijk} \sim N(0, \sigma_e^2)$
- We are interested in testing

$$H_0 : \tau \leq 0, \qquad H_1 : \tau > 0.$$

- We want a type-I error-rate of α when $\tau = 0$, and power of $1 - \beta$ when $\tau = \delta$

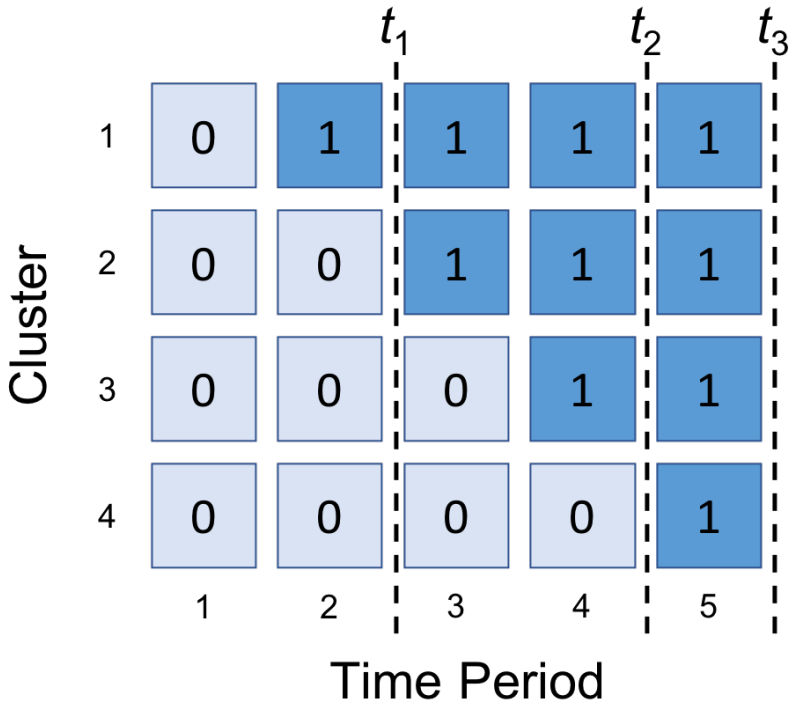
Example

- Bashour *et al* (2013) described a SW-CRT to assess the effect of training doctors in communication skills on women's satisfaction with doctor-woman relationship during labour and delivery
- Design had $C = 4$, $T = 5$, estimated $\hat{\tau} = -0.13$, $\hat{\sigma}_c^2 = 0.02$ and $\hat{\sigma}_e^2 = 0.51$
- For $\alpha = 0.05$, $\beta = 0.1$, $\delta = 0.2$, design required $m = 70$ patients per cluster per period
- Use these as the example parameters in all of what follows

Outline

Goal

- Include interim analyses where we can test for futility/efficacy, reducing the required number of observations
- Specify a collection of time periods after which we will conduct analyses: $t_1, t_2, \ldots$



Stopping rules

- Our testing rules are then based on efficacy and futility stopping boundaries, $\boldsymbol{e} = (e_{t_1}, e_{t_2}, \dots)$ and $\boldsymbol{f} = (f_{t_1}, f_{t_2}, \dots)$
- We need a test statistic Z_t to use after time period t
  - If $Z_t > e_t$ then we stop the trial, and reject H_0
  - If $Z_t \leq f_t$ then we stop the trial, and do not reject H_0
  - Otherwise we continue the trial to period $t + 1$
- Natural to use a Wald test statistic

$$Z_t = \frac{\hat{\tau}_t}{\sqrt{\text{Var}(\hat{\tau}_t)}} = \hat{\tau}_t I_t^{1/2}.$$

Theory

- Established group sequential design methodology is directly applicable to this longitudinal setting
- $\boldsymbol{Z} = (Z_{t_1}, Z_{t_2}, \dots)$ has a multivariate normal distribution

$$\mathbb{E}(Z_t) = \tau I_t^{1/2},$$
$$\text{Cov}(Z_{t_i}, Z_{t_i}) = (I_{t_i}/I_{t_j})^{1/2}, \ t_i \leq t_j.$$

- So we can compute the probability we stop for efficacy/futility at each analysis using multivariate normal integration
- Adding up the probability you stop for efficacy at each analysis gives you the overall rejection probability
- Use this in...

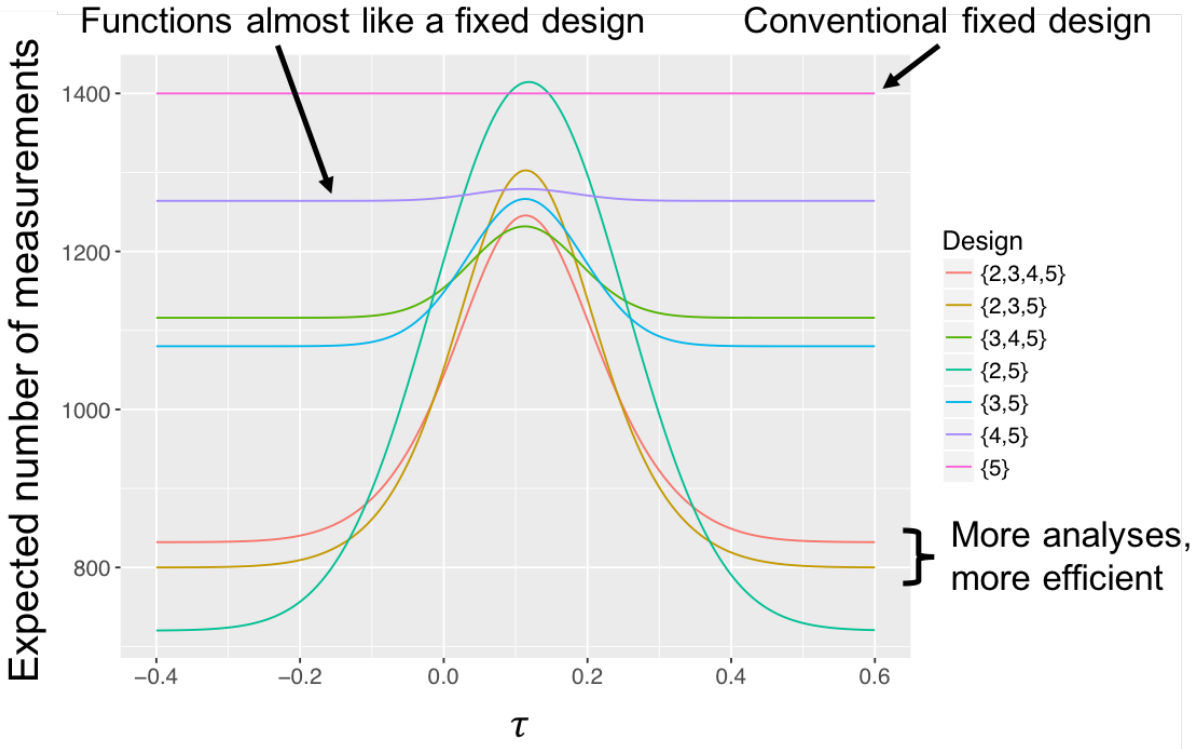
Design determination

- We need to choose m, $\boldsymbol{e}$, and $\boldsymbol{f}$, given choices of the variance components σ_c^2 and σ_e^2, that provide the desired error-rates
- A simple solution is to use the "error spending" approach to sequential trial design
- Or a global optimisation algorithm can be used to find the best possible boundaries
- Created some simple software for this

```
> example <- gs_sw(X, sigma_e, sigma_c, alpha,
                   beta, delta, set_T)
> summary(example)
  Identified design has:
    m = 97,
    e = (2.32, 2.06, 1.90),
    f = (-0.57, 0.65, 1.90).
```

Expected number of measurements

- Consider the influence of the choice of analysis times



- **But** including interim analyses increases the maximal possible required number of measurements

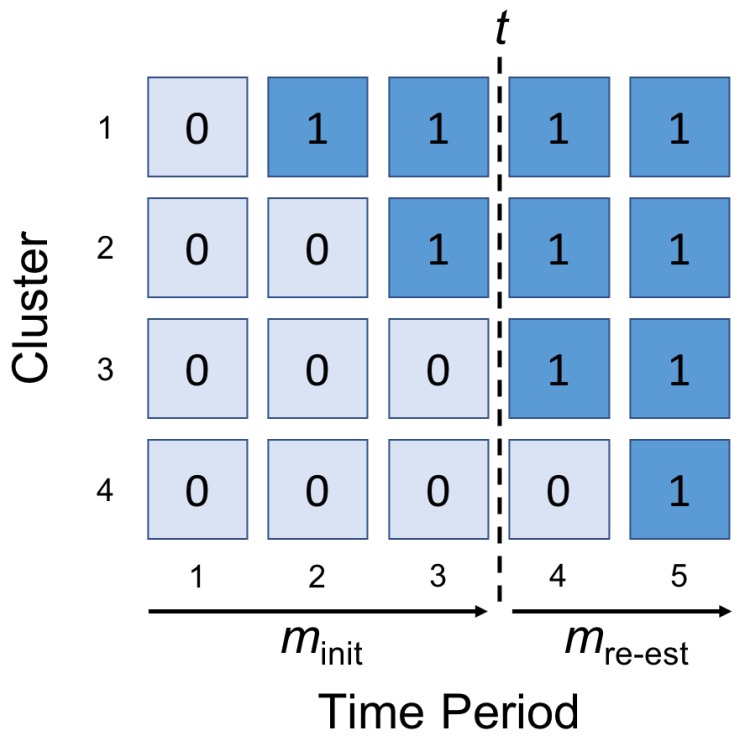
Summary

- Long-standing group sequential design theory can be used to incorporate interim analyses in to SW-CRTs
- Could be an effective method for reducing the expense of such trials when an intervention is highly effective/ineffective
- To power correctly, requires an assumption of known variance...

Outline

Goal

- After some specified time period t, re-estimate the required variance components
- Go from m_{init} measurements per cluster per period to a hopefully more appropriate $m_{\text{re-est}}$



Unblinded re-estimation

- Always straight forward to implement, but less favoured by regulatory agencies
- Fit the chosen model to the data to acquire the re-estimates $\hat{\sigma}_c^2$ and $\hat{\sigma}_e^2$
- Can write down a variance for the information when m_{init} patients are used in time periods $1, \ldots, t$, and $m_{\text{re-est}}$ in time periods $t+1, \ldots, T$

$$\text{Var}(\hat{\tau} \mid m_{\text{init}}, m_{\text{re-est}}, \hat{\sigma}_c^2, \hat{\sigma}_e^2).$$

- Use these in the conventional method of sample size determination

$$\Phi\{\delta/\text{Var}(\hat{\tau} \mid m_{\text{init}}, m_{\text{re-est}}, \hat{\sigma}_c^2, \hat{\sigma}_e^2) - z_{1-\alpha}\} \geq 1 - \beta.$$

Blinded re-estimation

- Works in a similar way, but to acquire out estimates we use

$$S^2_{\text{eq1}} = \frac{m_{\text{init}}}{Ct - t} \sum_{i=1}^{C} \sum_{j=1}^{t} (\bar{Y}_{ij.} - \bar{Y}_{.j.})^2,$$

$$S^2_{\text{eq2}} = \frac{1}{m_{\text{init}} Ct - Ct} \sum_{i=1}^{C} \sum_{j=1}^{t} \sum_{k=1}^{m_{\text{init}}} (Y_{ijk} - \bar{Y}_{ij.})^2.$$

- We can show that in the absence of a treatment effect

$$\hat{\sigma}^2_e = S^2_{\text{eq2}},$$

$$\hat{\sigma}^2_c = \frac{1}{n}(S^2_{\text{eq1}} - S^2_{\text{eq2}}),$$

are unbiased estimates of the variance components

Simulation study: Power

Assumed variances	Blinded	Unblinded	Fixed
50% smaller than truth	0.881	0.880	0.692
Correct	0.885	0.884	0.903
50% larger than truth	0.897	0.895	0.974

Procedure	Assumed variances	$t = 2$	$t = 3$	$t = 4$
Blinded	50% smaller than truth	0.870	**0.881**	**0.815**
Blinded	Correct	0.885	0.885	0.897
Blinded	50% larger than truth	0.888	**0.897**	**0.954**
Unblinded	50% smaller than truth	0.869	0.880	0.817
Unblinded	Correct	0.884	0.884	0.897
Unblinded	50% larger than truth	0.887	0.896	0.955

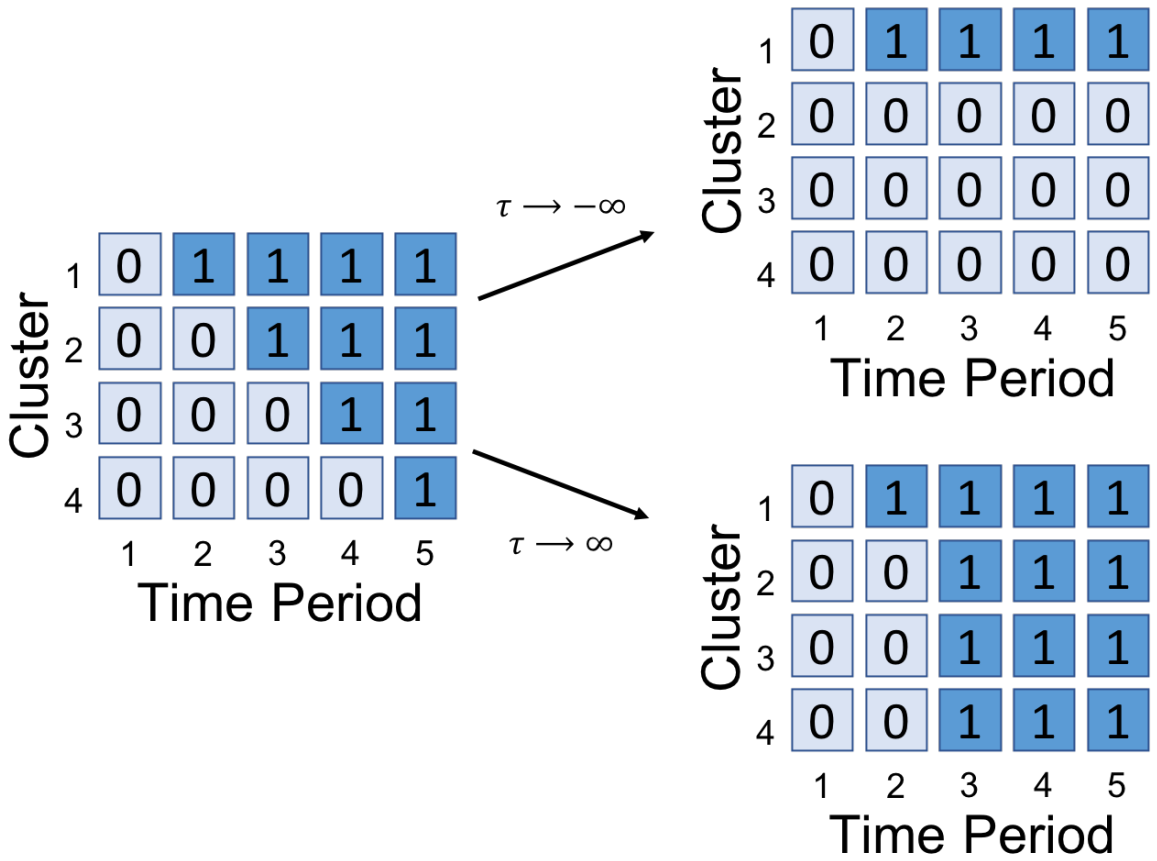
Summary

- Sample size re-estimation can greatly increase your power when you have under-specified the variance components
- Only a single re-estimation point required. Feasible in practice?
- Need to think carefully about when to time the re-estimation
- One other issue is that it is quite computationally intensive to investigate these designs

Outline

Discussion

- A variety of adaptive designs are available for CRTs
- Also working on a technique for response adaptive treatment allocation



Discussion

- Time period structure of SW-CRTs may make interim analyses particularly appealing
- However, there are issues associated with their use
- Can interim analyses be handled efficiently?
- If you stop for efficacy, what would you do you do next if resources are scarce?
- Can you realistically handle an increase in the cluster period sample size?
- Nonetheless, if used wisely they could help greatly with improving efficiency/power of SW-CRTs

References

- Bashour HN, Kanaan M, Kharouf MH, Abdulsalam AA, Tabbaa MA, Cheikha SA (2013). The effect of training doctors in communication skills on women's satisfaction with doctor-woman relationship during labour and delivery: a stepped wedge cluster randomised trial in Damascus. *BMJ Open* **3**:e002674
- Grayling MJ, Wason JMS, Mander AP (2017) Group sequential designs for stepped-wedge cluster randomised trials. *Clin Trials* **14:**507-17
- Grayling MJ, Mander AP, Wason JMS (2018) Blinded and unblinded sample size reestimation procedures for stepped-wedge cluster randomized trials. *Biom J* **60:**903-16
- Grayling MJ, Robertson DS, Wason JMS, Mander AP (2018) Design optimisation and post-trial analysis in group sequential stepped-wedge cluster randomised trials. arXiv:1803.09691v1
- Lake S, Kammann E, Klar N, Betensky R (2002) Sample size re-estimation in cluster randomization trials. *Stat Med* **21:**1337–50
- van Schie S, Moerbeek M (2014) Re-estimating sample size in cluster randomised trials with active recruitment within clusters. *Stat Med* **33:**3253–68
- Zou GY, Donner A, Klar N (2005) Group sequential methods for cluster randomization trials with binary outcomes. *Clin Trials* **2:**479-87