Comparative Efficiency of Different Cluster Designs in Practice

Jessica Kasza

jessica.kasza@monash.edu

School of Public Health and Preventive Medicine
Monash University

Choosing the right design

Design choice depends on context:

- Number of available clusters (K);
- Number of available participants within each cluster (m);
- Length of time available to run trial;
- Whether implementation or wash-out periods are necessary;
- Whether it is possible to switch back and forth between treatments.

Many different designs possible for any given setting!

Choosing the right design

Design choice depends on context:

- Number of available clusters (K);
- Number of available participants within each cluster (m);
- Length of time available to run trial;
- Whether implementation or wash-out periods are necessary;
- Whether it is possible to switch back and forth between treatments.

Many different designs possible for any given setting!

How do we choose?

Choosing a design

Generally, want a design that:

- Satisfies constraints (number of clusters, participants, etc...);
- Can feasibly be implemented;
- Minimises risk of bias (not discussed today - come back tomorrow!);
- Has enough (i.e. at least 80%) statistical power to detect the effect of interest.

Choosing a design

Generally, want a design that:

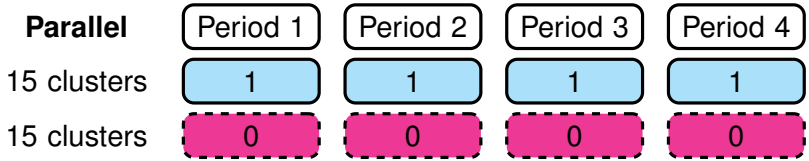
- Satisfies constraints (number of clusters, participants, etc...);
- Can feasibly be implemented;
- Minimises risk of bias (not discussed today - come back tomorrow!);
- Has enough (i.e. at least 80%) statistical power to detect the effect of interest.
  - The power of a study depends on the study design through the variance of the treatment effect estimator
  - A more efficient design (statistically speaking) is that with a lower treatment effect estimator variance, requiring fewer participants.

The ICU example

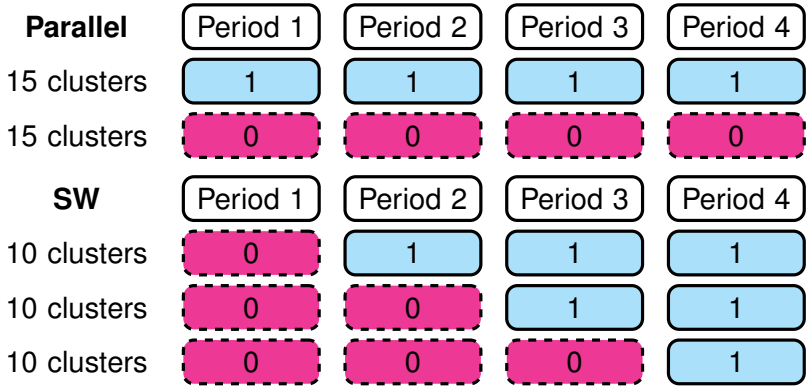
Does overnight placement of earplugs in patients in the ICU reduce hospital length of stay?

- 4-period cluster trial with 30 clusters:
  - Consider parallel, stepped wedge, and CRXO designs.

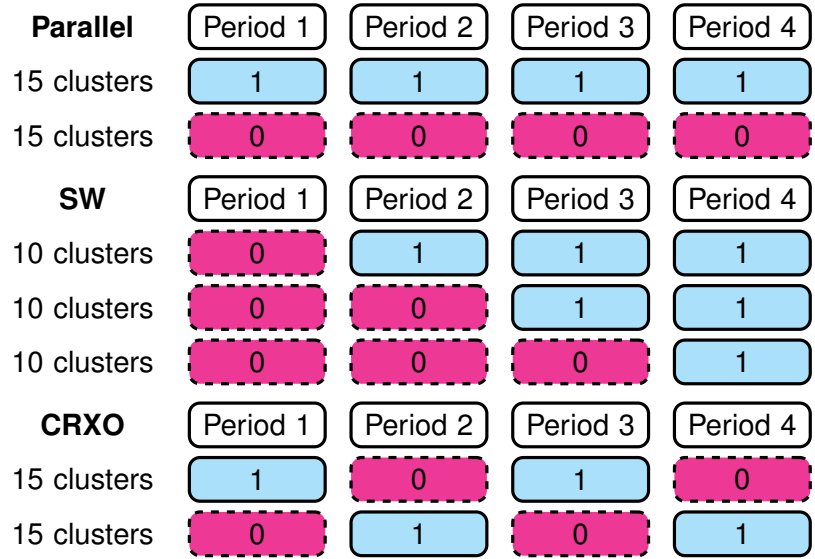
ICU example considered designs



ICU example considered designs



ICU example considered designs



Estimating correlation parameters

- A fortunate situation!
  - A large database is available: we use this to investigate the within-cluster correlation structure.

Assume a model for the logarithm of hospital length-of-stay:

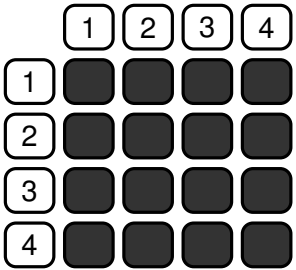
$$\log(\text{Hosp. LoS}) = \text{Period effect} + \text{Treatment effect} + \textbf{random effects} + \text{error}$$

Random effects:

- Encode the similarity of patients from the same ICU.
- Many different assumptions about these random effects can be made!
- We considered three different models for the random effects.

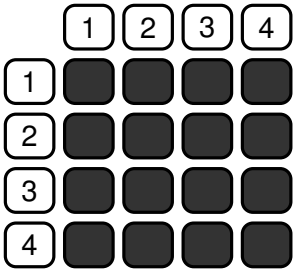
Within- and between-period ICCs, 4-period design

Model 1:
no decay over time

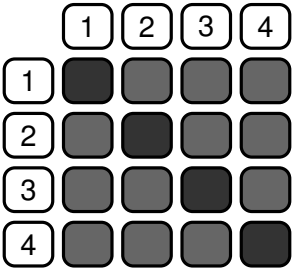


Within- and between-period ICCs, 4-period design

Model 1:
no decay over time

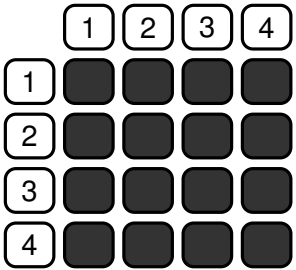


Model 2: within ICC $\neq$ between ICC

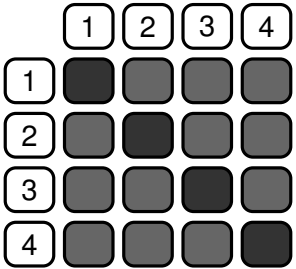


Within- and between-period ICCs, 4-period design

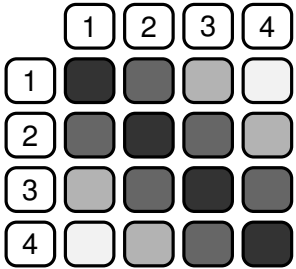
Model 1:
no decay over time



Model 2: within ICC $\neq$ between ICC



Model 3: ICC that decays over time



Within-period and between-period ICCs

Model	Within-period ICC $corr(Y_{kti}, Y_{ktj})$	Between-period ICC $corr(Y_{kti}, Y_{ksj}), s \neq t$
1	ρ	ρ
2	ρ	$r \times \rho$
3	ρ	$r^{\lvert t-s \rvert} \times \rho$

Within-period and between-period ICCs

Model	Within-period ICC $corr(Y_{kti}, Y_{ktj})$	Between-period ICC $corr(Y_{kti}, Y_{ksj}), s \neq t$
1	ρ	ρ
2	ρ	$r \times \rho$
3	ρ	$r^{\lvert t-s \rvert} \times \rho$

For Model 3:

- Intracluster correlation: $\hat{\rho} = 0.035$
- Decay per period: 5% $\Rightarrow$ Cluster AutoCorrelation 0.95.
  - Correlation between two patients in the same ICU in the same period: 0.035
  - Correlation between two patients in the same ICU in periods 1 and 2: $0.035 \times 0.95 = 0.033$
  - Correlation between two patients in the same ICU in periods 1 and 3: $0.035 \times 0.95^2 = 0.031$

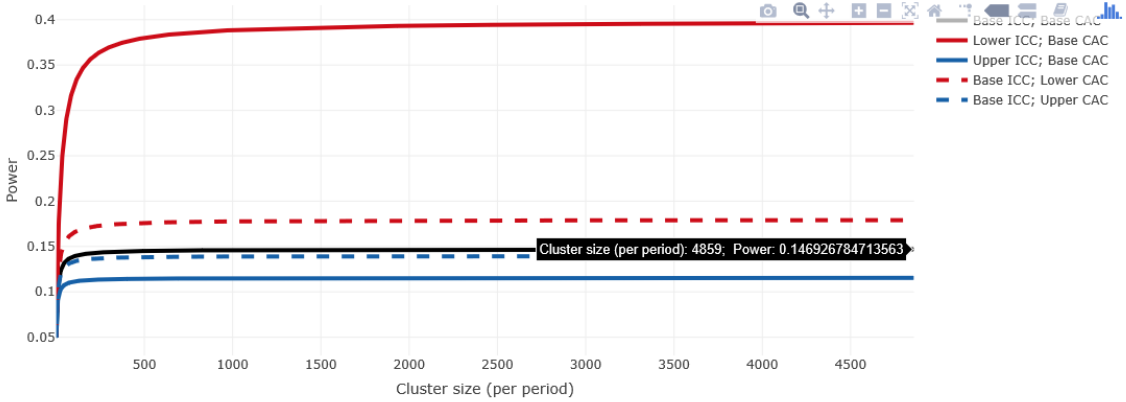
Task 1

- Cross-sectional sampling structure
- Within-period ICC (ρ): 0.035.
- Cluster auto-correlation: 0.95, and allow for decaying CAC over time.
- Effect size 0.06, 80% power, 5% level of significance.

What is the minimum number of subjects per cluster per period to detect this effect size for the parallel, SW, and CRXO designs?

Task 1: parallel

Power | Precision | Design matrix | References and Contacts



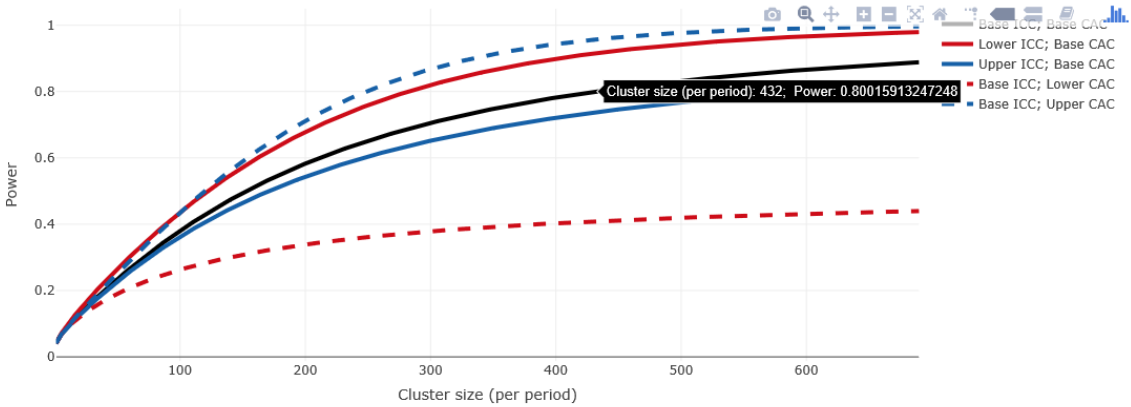
Curve shows the increase in power as the cluster-period size increases (for a fixed number of clusters). Hover cursor over curve to see actual power values

Warning: caution is needed with CRTs with a small number of clusters due to risk of lack of internal and external validity; and appropriateness of calculations used particularly for binary and count outcomes

Parameters:

Task 1: stepped wedge

Power | Precision | Design matrix | References and Contacts



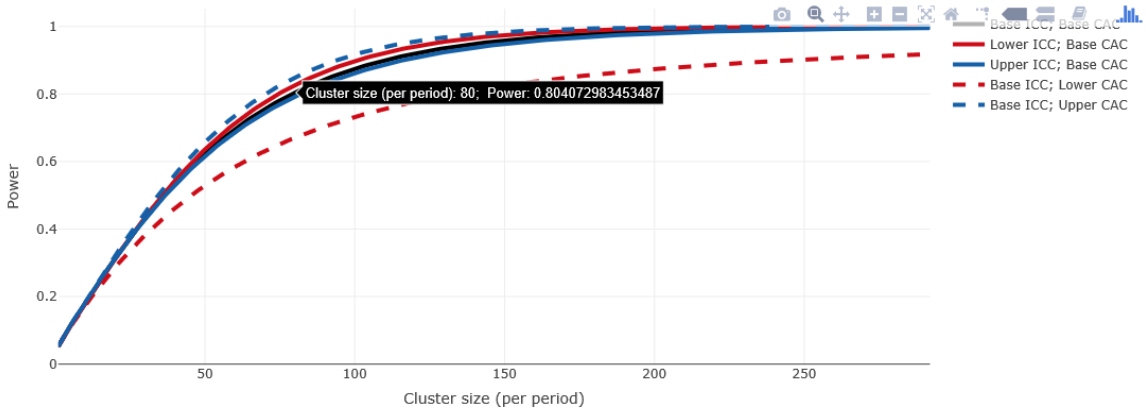
Curve shows the increase in power as the cluster-period size increases (for a fixed number of clusters). Hover cursor over curve to see actual power values

Warning: caution is needed with CRTs with a small number of clusters due to risk of lack of internal and external validity; and appropriateness of calculations used particularly for binary and count outcomes

Parameters:

Task 1: multi cross-over

Power | Precision | Design matrix | References and Contacts



Curve shows the increase in power as the cluster-period size increases (for a fixed number of clusters). Hover cursor over curve to see actual power values

Warning: caution is needed with CRTs with a small number of clusters due to risk of lack of internal and external validity; and appropriateness of calculations used particularly for binary and count outcomes

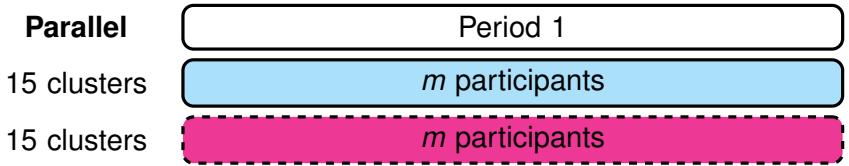
Task 1: comparing parallel, SW, and CRXO

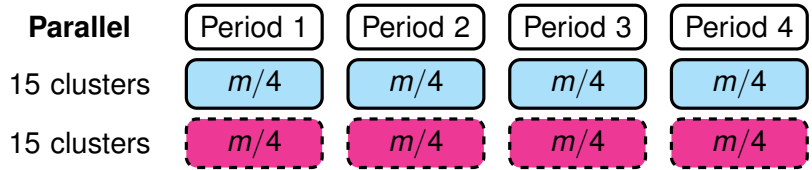
- For the parallel design, with this number of clusters and within-cluster correlation structure, doesn't matter how many patients we recruit in each ICU - won't achieve required power.
- Total number of patients for the SW: $30 \times 4 \times 432 = 51,840$
- Total number of patients for the CRXO: $30 \times 4 \times 80 = 9,600$

Crossing back and forth leads to very large sample size savings!

An aside: the multiple-period parallel design

Why did we consider the multiple period parallel design instead of just the parallel design?





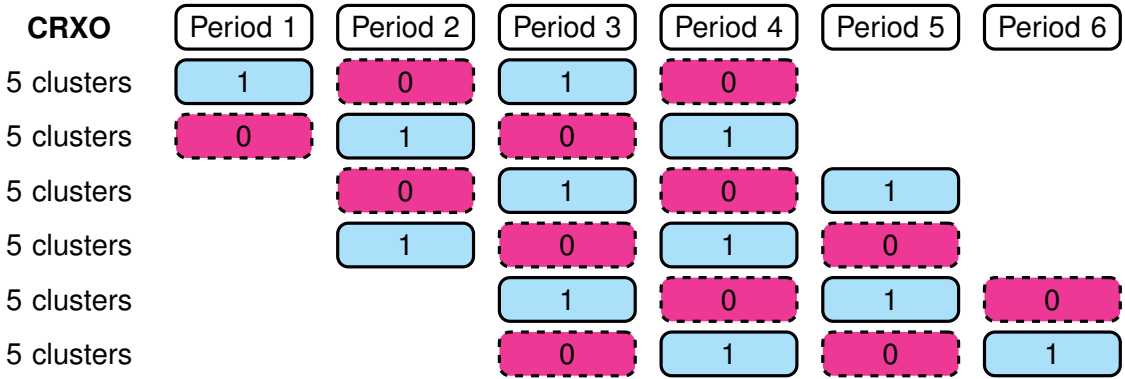
An aside: the multiple-period parallel design

Why did we consider the multiple period parallel design instead of just the parallel design?

- The single-period parallel design assumes that all participants have outcomes that are **equally correlated**.
  - For the ICU example, we considered the discrete-time decay model.
- For the comparison between the parallel, SW and CRXO designs to make sense, need consistency across correlation structures.

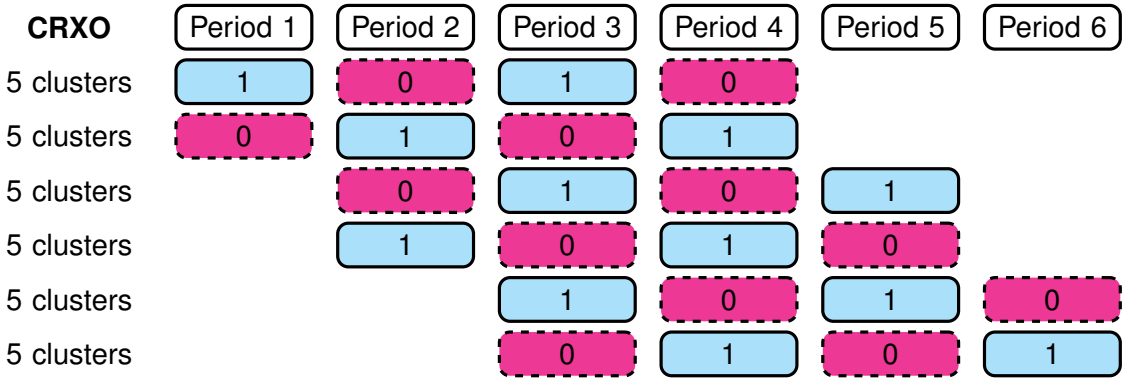
Task 2: varying cluster start time

What if not all clusters can start at the same time? What happens to the number of patients required in each cluster-period?



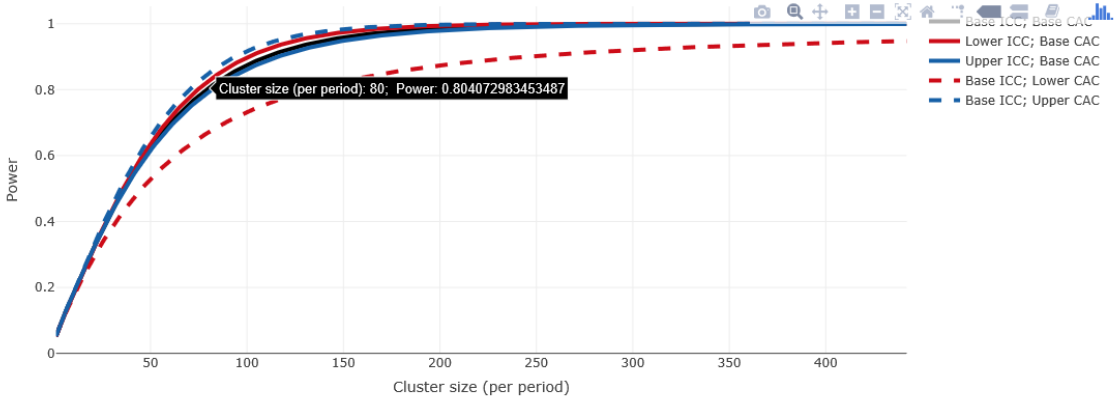
Task 2: varying cluster start time

What if not all clusters can start at the same time? What happens to the number of patients required in each cluster-period?



Task 2: varying cluster start time

Power | Precision | Design matrix | References and Contacts



Curve shows the increase in power as the cluster-period size increases (for a fixed number of clusters). Hover cursor over curve to see actual power values

Warning: caution is needed with CRTs with a small number of clusters due to risk of lack of internal and external validity; and appropriateness of calculations used particularly for binary and count outcomes

Parameters:

Task 2: varying cluster start time

For the CRXO:

- As long as the design is **balanced** (equal numbers of clusters in both treatment groups at each period) then delays to start won't change study power.

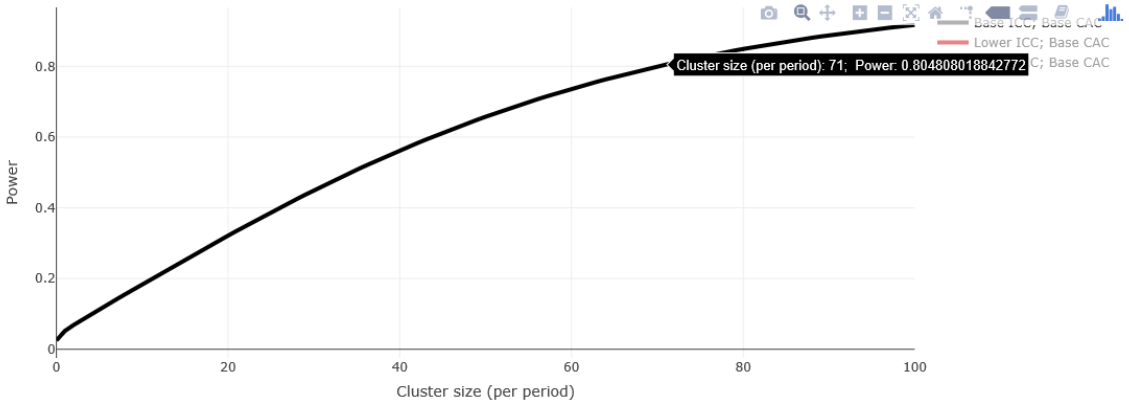
Task 3: omitting decays in correlations

- Cross-sectional sampling structure.
- Within-period ICC (ρ): 0.035.
- Cluster auto-correlation: 0.95, and allow for decaying CAC over time.

What happens to the required number of patients in each cluster in each period when CAC =1 instead of 0.95 with a decay, for the CRXO and SW designs?

Task 3: CRXO with no decay

Power | Precision | Design matrix | References and Contacts



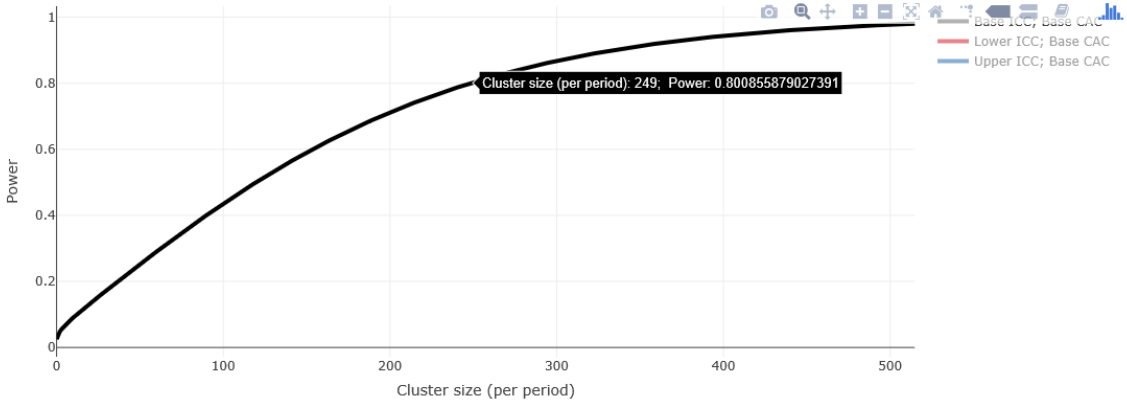
Curve shows the increase in power as the cluster-period size increases (for a fixed number of clusters). Hover cursor over curve to see actual power values

Warning: caution is needed with CRTs with a small number of clusters due to risk of lack of internal and external validity; and appropriateness of calculations used particularly for binary and count outcomes

Parameters:

Task 3: SW with no decay

Power | Precision | Design matrix | References and Contacts



Curve shows the increase in power as the cluster-period size increases (for a fixed number of clusters). Hover cursor over curve to see actual power values

Warning: caution is needed with CRTs with a small number of clusters due to risk of lack of internal and external validity; and appropriateness of calculations used particularly for binary and count outcomes

Parameters:

Task 3

- CRXO design: 80 participants in each cluster period with decay; 71 without.
- SW design: 432 participants in each cluster period with decay; 249 without.

If a decay in the correlation is ignored **at the design stage**, end up with a sample size 1.1 (CRXO) and 1.7 (SW) times too small!

Task 3

- CRXO design: 80 participants in each cluster period with decay; 71 without.
- SW design: 432 participants in each cluster period with decay; 249 without.

If a decay in the correlation is ignored **at the design stage**, end up with a sample size 1.1 (CRXO) and 1.7 (SW) times too small!

Why is the sample size increased when a decay is included?

- The CRXO and SW capitalise on within-cluster comparisons.
  - The more similar participants in a cluster are, the more benefit there is from comparing within a cluster!
- When a decay is included, the similarity of participants declines as they move further apart in time.

Task 3

- CRXO design: 80 participants in each cluster period with decay; 71 without.
- SW design: 432 participants in each cluster period with decay; 249 without.

If a decay in the correlation is ignored **at the design stage**, end up with a sample size 1.1 (CRXO) and 1.7 (SW) times too small!

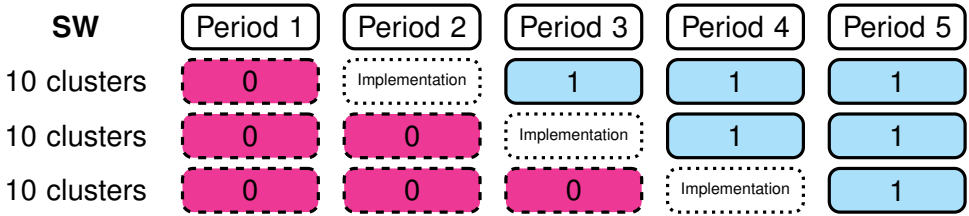
Why is the sample size increased when a decay is included?

- The CRXO and SW capitalise on within-cluster comparisons.
  - The more similar participants in a cluster are, the more benefit there is from comparing within a cluster!
- When a decay is included, the similarity of participants declines as they move further apart in time.

Follow-up questions:

- The CRXO seems to be less sensitive to the omission of the decay - why?
- What if a decay is ignored when analysing outcomes?

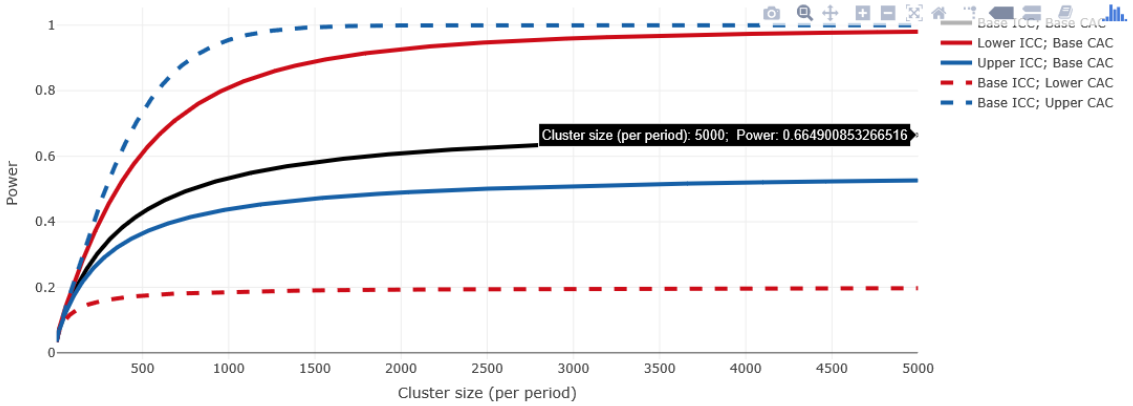
Task 4: Including an implementation period in the SW design



For the ICU example (ICC=0.035, decay = 0.95, effect size = 0.06, 80% power, 5% level of significance)

Task 4: SW with implementation periods

Power | Precision | Design matrix | References and Contacts



Curve shows the increase in power as the cluster-period size increases (for a fixed number of clusters). Hover cursor over curve to see actual power values

Warning: caution is needed with CRTs with a small number of clusters due to risk of lack of internal and external validity; and appropriateness of calculations used particularly for binary and count outcomes

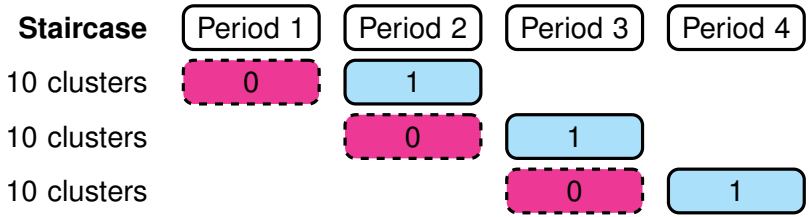
Parameters:

Task 4: SW with implementation periods

- Even with 5,000 patients in each cluster in each period, still don't get 80% power!
- Introducing an implementation period causes problems for this design
  - Direct comparisons between clusters in only one period!

Task 5: incomplete SW designs

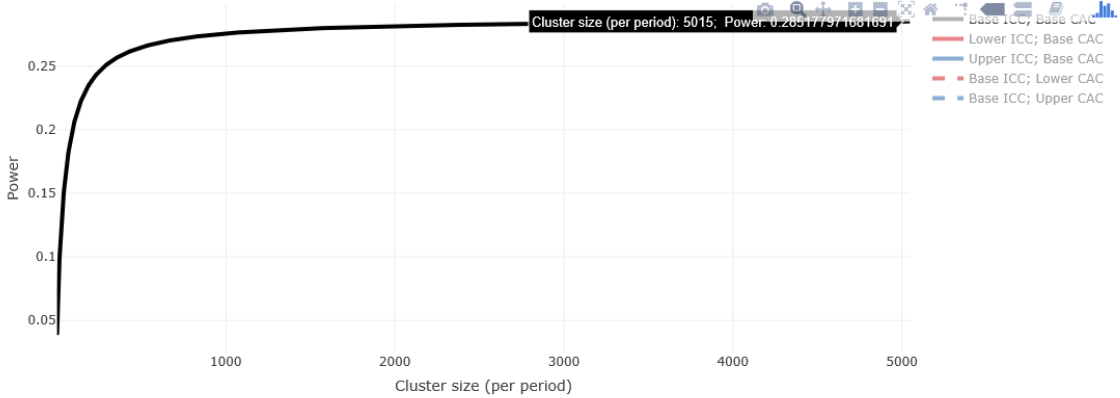
- Some recent work has indicated that in the stepped wedge design, the cluster-periods near the time of the treatment switch contain a lot of the information about the treatment effect.
- So we might consider a staircase design:



- How many participants per cluster-period are required to achieve 80% power for the staircase design for the ICU example?

Task 5: incomplete SW designs

Power | Precision | Design matrix | References and Contacts



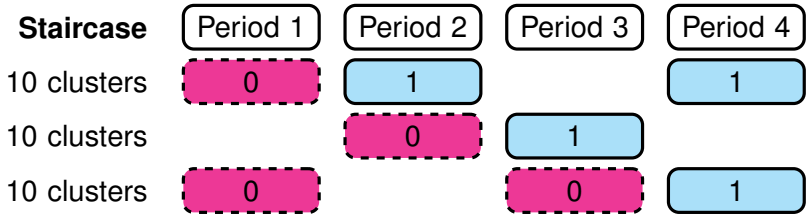
Curve shows the increase in power as the cluster-period size increases (for a fixed number of clusters). Hover cursor over curve to see actual power values

Warning: caution is needed with CRTs with a small number of clusters due to risk of lack of internal and external validity; and appropriateness of calculations used particularly for binary and count outcomes

Parameters:

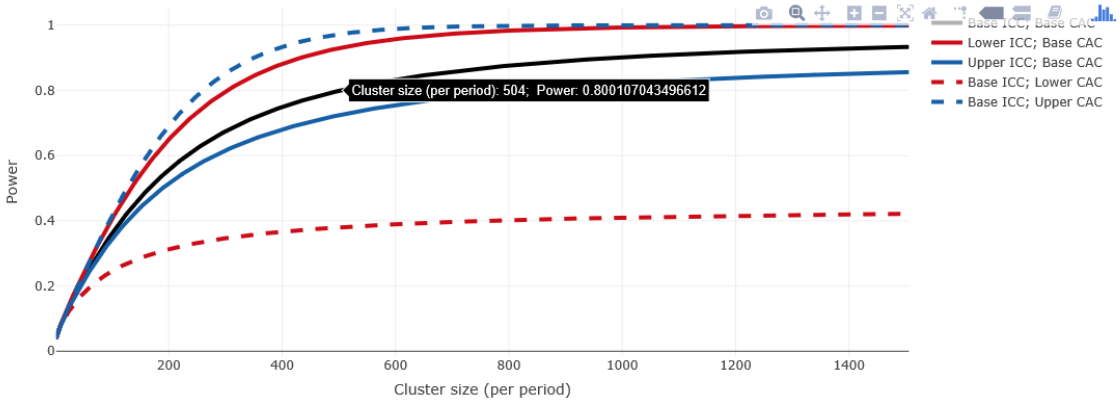
Task 5: incomplete SW designs

- What about for the staircase + corner design:



Task 5: incomplete SW designs

Power | Precision | Design matrix | References and Contacts



Curve shows the increase in power as the cluster-period size increases (for a fixed number of clusters). Hover cursor over curve to see actual power values

Warning: caution is needed with CRTs with a small number of clusters due to risk of lack of internal and external validity; and appropriateness of calculations used particularly for binary and count outcomes

Parameters:

Task 5: incomplete SW designs

- Lose a lot of information when only consider the staircase design!
- ... but incomplete designs such as this may be appealing to clusters and participants.
- If the “corners” are also included, get back a lot of information.

Key points

- When comparing different designs, make sure to compare like with like:
  - The within-cluster correlation structures must be consistent across the compared designs.
  - The estimated correlations should be estimated using the same period length as in the planned study.
- For designs with treatment switching (e.g. stepped wedge and multi cross-over) the impact of exponentially decaying correlations on required sample sizes can be large.
- Important for sample size calculation to reflect the trial as it will be conducted as closely as possible.
- The length of the design periods needs to be considered: if the number of participants that can be included/recruited in one period is limited, the length of periods may need to be adjusted (may have implications for the estimated correlations!)