A talk in two halves:
1. Some recent work in longitudinal cluster randomised trials
2. Increasing the representation of women in statistics

Jessica Kasza, @JessKasza

jessica.kasza@monash.edu

School of Public Health and Preventive Medicine
Monash University

Outline

Outline

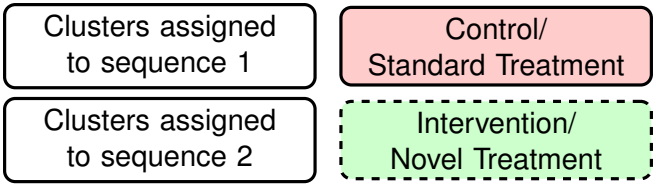
How can we work out if an intervention works?

- We have a new treatment that we hope reduces the severity of COVID-19 infections. How can we tell if it works?
- Gold standard: randomise patients with COVID-19 to receive the new treatment, or to receive the standard treatment.
  - By randomisation, if the trial is well-conducted, any differences between groups due to the treatment.

How can we work out if an intervention works?

- We have a new treatment that we hope reduces the severity of COVID-19 infections. How can we tell if it works?
- Gold standard: randomise patients with COVID-19 to receive the new treatment, or to receive the standard treatment.
  - By randomisation, if the trial is well-conducted, any differences between groups due to the treatment.
- We have a new hospital cleaning program. Does it reduce the number of hospital-acquired infections?
- The intervention is applied to hospitals (not to the individual patients).
  - So how can we randomise individual patients?

The standard cluster randomised trial



- *Clusters (groups)* of participants assigned to treatments.
  - Necessary when treatment applied at the group level;
  - Useful when there may be contamination between participants;
  - Perhaps an individually randomised trial would be logistically infeasible (Beware ethical issues!).
- Clusters could be hospitals, ICUs, schools, neighbourhoods...
  - The **intracluster correlation (ICC)** (ρ) describes the similarity of outcomes from participants in the same cluster.

Extending cluster randomised trials over time

Cluster randomised trials not always feasible:

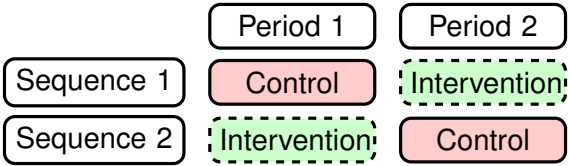
- The ICC ρ reduces the amount of information we can get from each subject.
- What if more clusters are required than are available?

Longitudinal cluster randomised trials can help!

- Clusters are followed over time and measured repeatedly.
  - Clusters may switch between control and intervention conditions.
  - Subjects could be measured in one period only, in all periods, or in some other number of periods.

The cluster randomised cross over trial: CRXO

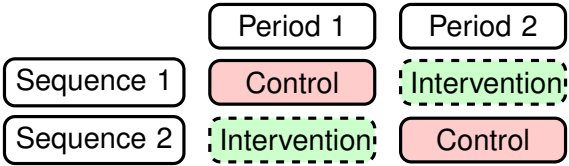
The standard 2-period CRXO:



- Each cluster switches: treatment effect is now estimated using **BOTH** within-cluster and between-cluster comparisons.

The cluster randomised cross over trial: CRXO

The standard 2-period CRXO:



- Each cluster switches: treatment effect is now estimated using **BOTH** within-cluster and between-cluster comparisons.

Now that we're following clusters up over time, need to think more about the ICC ρ...

CRXO: the cluster randomised crossover trial

Now need to consider two intracluster correlations!

- **Within-period ICC**: correlation between observations on subjects in the same cluster

in the same period

- **Between-period ICC**: correlation between observations on subjects in same cluster

in different periods

Would we expect the within-period and between-period ICCs to be identical?

CRXO: the cluster randomised crossover trial

Now need to consider two intracluster correlations!

- **Within-period ICC**: correlation between observations on subjects in the same cluster

in the same period

- **Between-period ICC**: correlation between observations on subjects in same cluster

in different periods

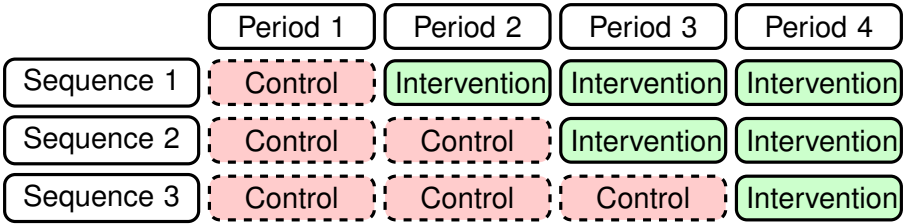
Would we expect the within-period and between-period ICCs to be identical?

$$\text{Within-period ICC} = \rho$$

$$\text{Between-period ICC} = r \times \rho$$

$r = \frac{\text{Between-period ICC}}{\text{Within-period ICC}}$ is known as the **cluster autocorrelation** (**CAC**).

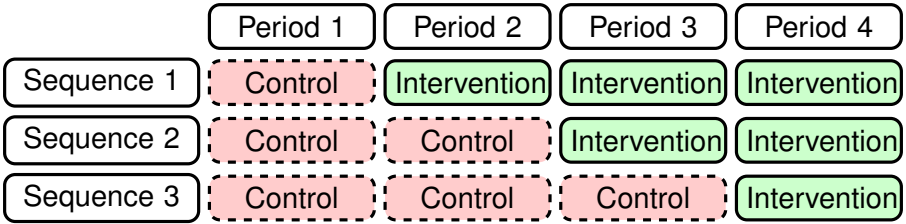
The stepped wedge cluster randomised trial design



Stepped wedge designs are wonderful!

- All clusters know they will receive the intervention (eventually...);
- Useful when interventions cannot be undone or will be rolled out anyway.

The stepped wedge cluster randomised trial design



Stepped wedge designs are wonderful!

- All clusters know they will receive the intervention (eventually...);
- Useful when interventions cannot be undone or will be rolled out anyway.

BUT... We now might expect that the between-period ICC *decays* over time.

Decaying between-period ICCs

- The correlation between participant outcomes decreases the further their measurement periods are apart in time.
  - For participants in period t and s: correlation is $\rho r^{|t-s|}$

Decaying between-period ICCs

- The correlation between participant outcomes decreases the further their measurement periods are apart in time.
  - For participants in period t and s: correlation is $\rho r^{|t-s|}$
- Intracluster correlation: $\rho = 0.035$
- Decay per period: 5% $\Rightarrow$ Cluster AutoCorrelation $r = 0.95$.
  - Correlation b/w two participants in the same cluster

in the same period: 0.035

in periods 1 and 2: $0.035 \times 0.95 = 0.033$

in periods 1 and 3: $0.035 \times 0.95^2 = 0.031$

in periods 1 and 4: $0.035 \times 0.95^3 = 0.030$

Outline

A comprehensive CRT sample size calculator

International Journal of Epidemiology, 2020, 1–17
doi: 10.1093/ije/dyz237
Original article

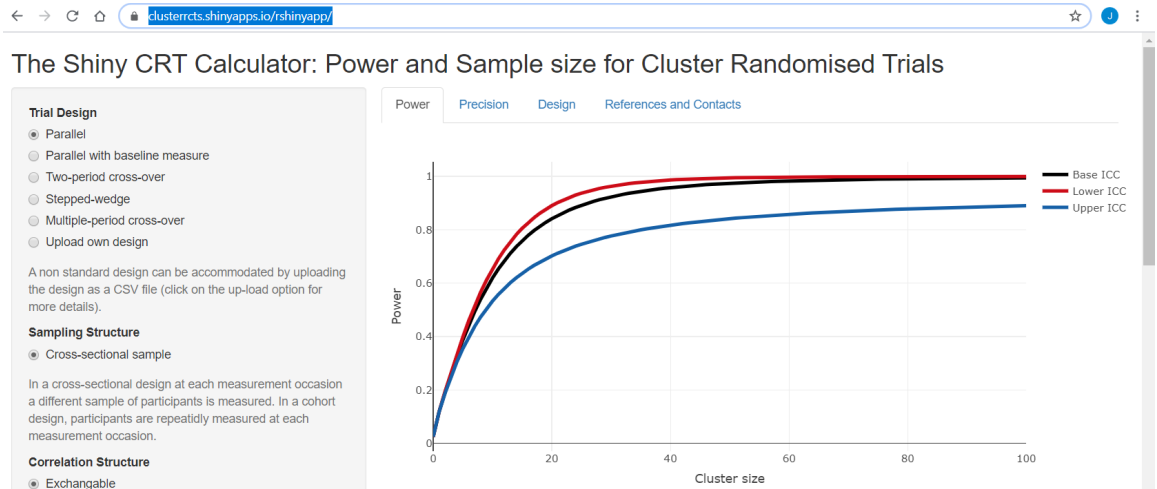
Original article

A tutorial on sample size calculation for multiple-period cluster randomized parallel, cross-over and stepped-wedge trials using the Shiny CRT Calculator

Karla Hemming [1]*, Jessica Kasza [2], Richard Hooper,[3] Andrew Forbes[2] and Monica Taljaard[4,5]

A comprehensive CRT sample size calculator

https://clusterrcts.shinyapps.io/rshinyapp/

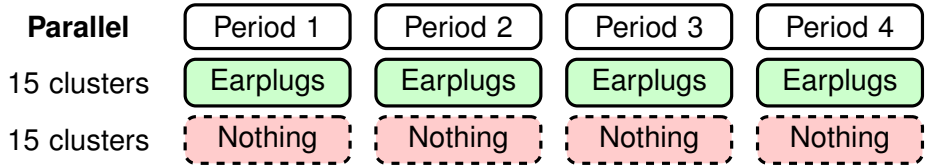


The Intensive Care Unit (ICU) example

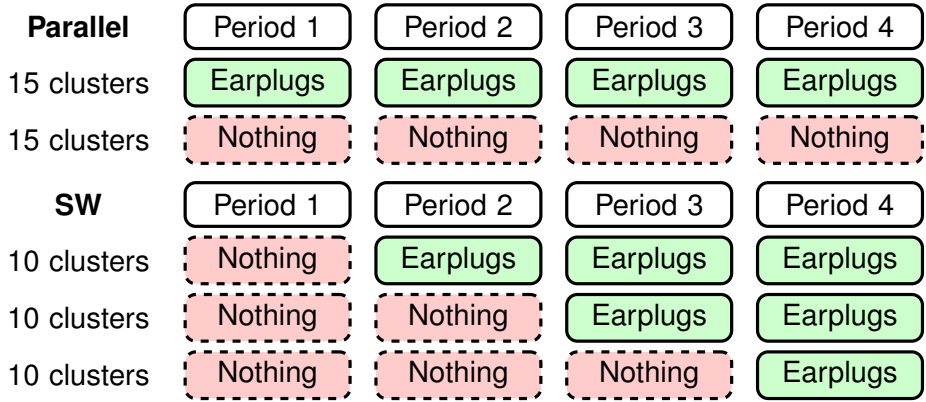
Does overnight placement of earplugs in patients in the ICU reduce hospital length of stay?

- About 30 ICUs are available in Australia, and we have about 2 years for the study.
  - Consider 4 six-month periods.
- 4-period cluster trial with 30 clusters:
  - Consider parallel, stepped wedge, and CRXO designs.

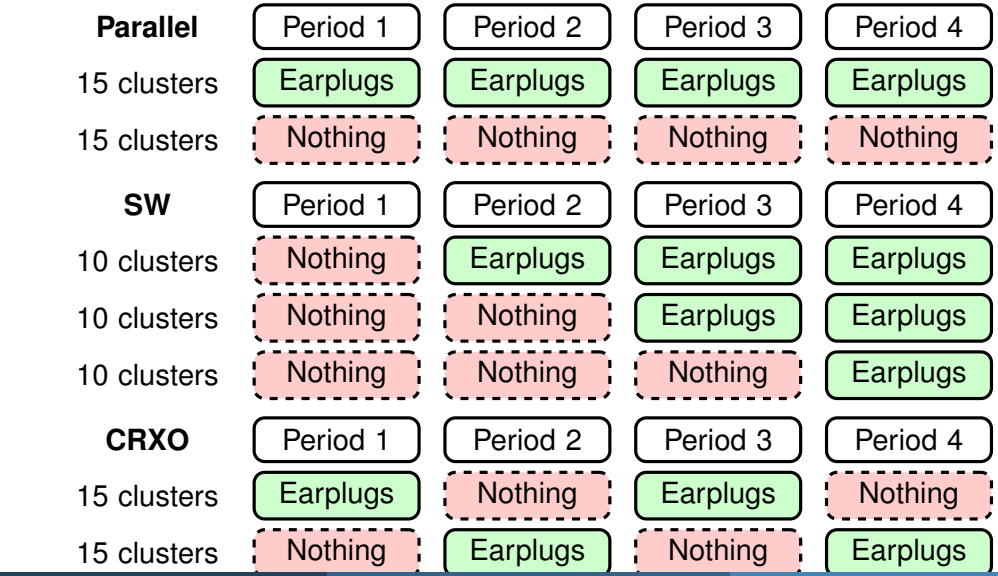
ICU example considered designs



ICU example considered designs



ICU example considered designs



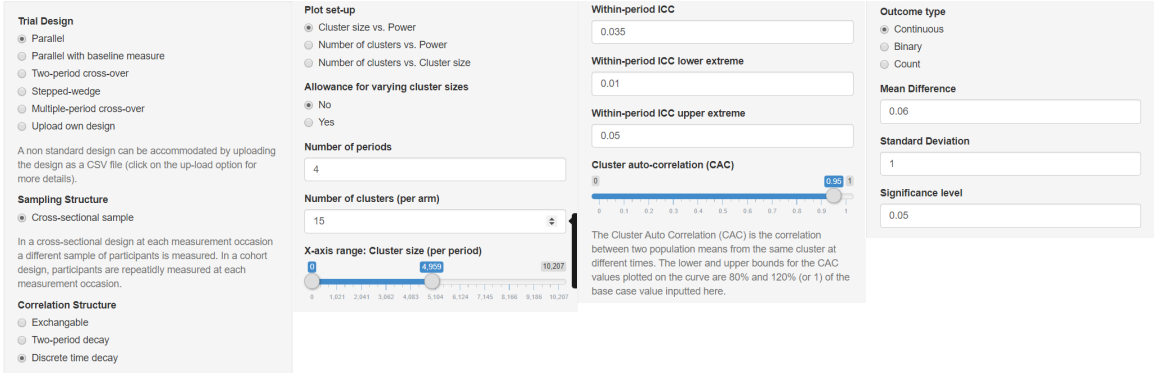
Which of the three designs requires fewest subjects for 80% power?

- Cross-sectional sampling structure
- Within-period ICC (ρ): 0.035.
- Cluster auto-correlation: $r = 0.95$, 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?

Parallel design

https://clusterrcts.shinyapps.io/rshinyapp/

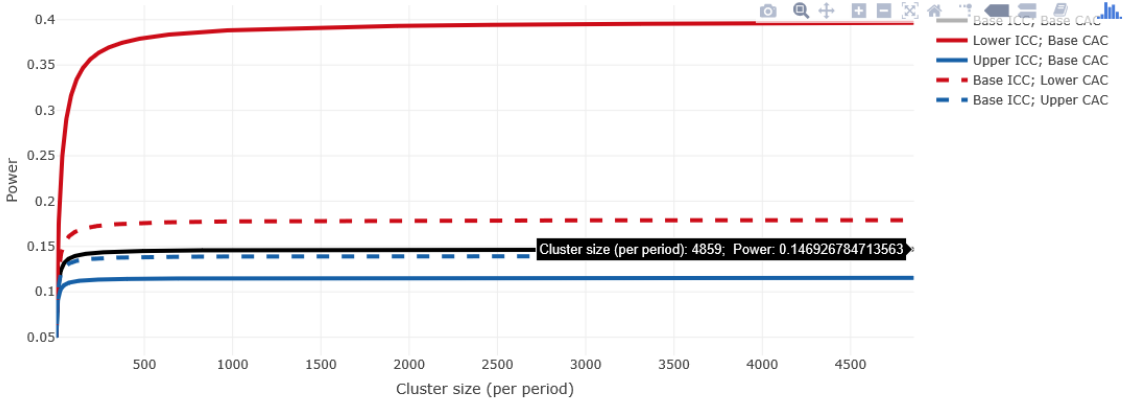


Parallel design

https://clusterrcts.shinyapps.io/rshinyapp/

Parallel design

Power | Precision | Design matrix | References and Contacts



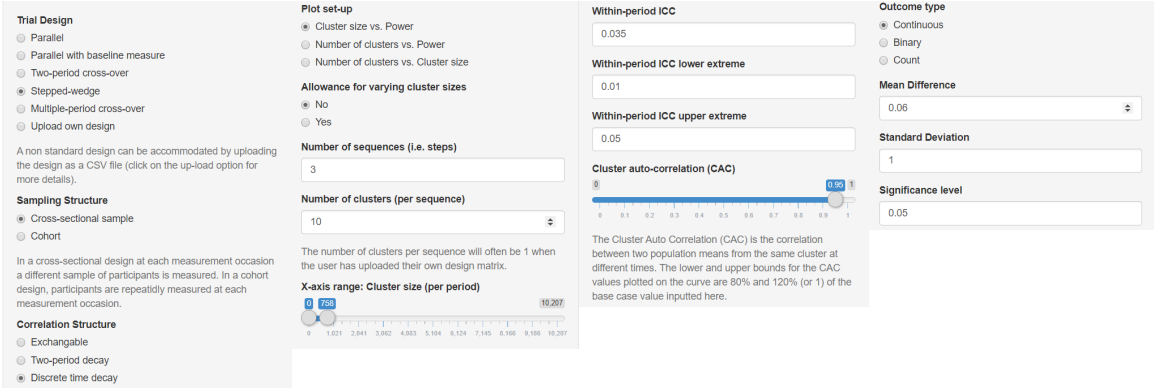
Curve shows the increase in power as the cluster-period size increases (for a fixed number of clusters). Hover cursor over curve to see actual power values

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:

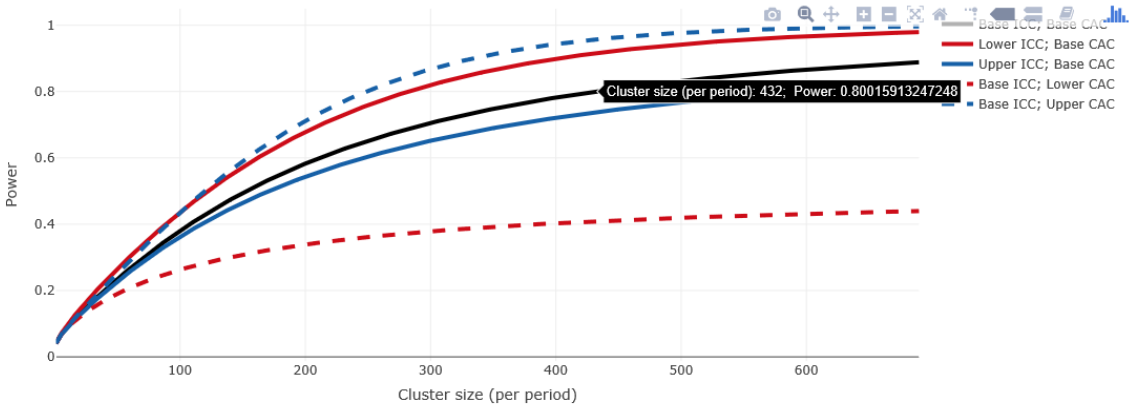
Stepped wedge

https://clusterrcts.shinyapps.io/rshinyapp/



Stepped wedge

Power | Precision | Design matrix | References and Contacts



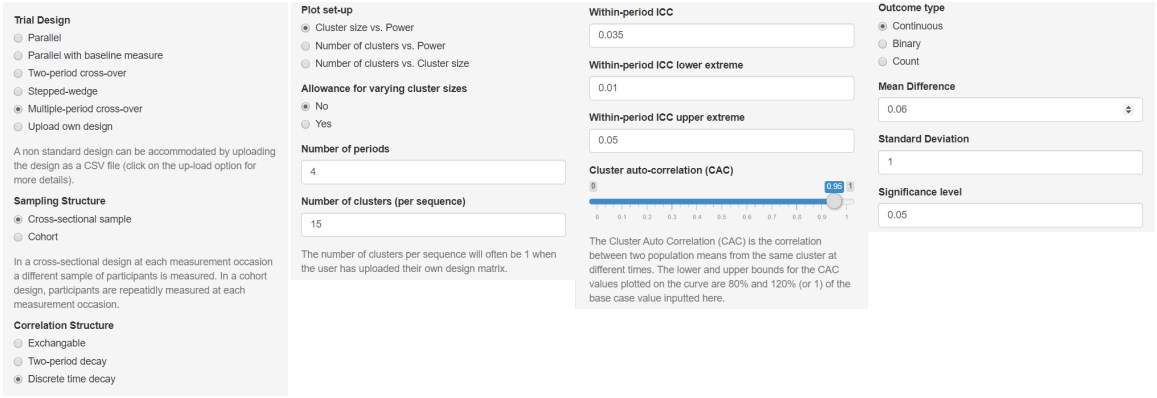
Curve shows the increase in power as the cluster-period size increases (for a fixed number of clusters). Hover cursor over curve to see actual power values

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:

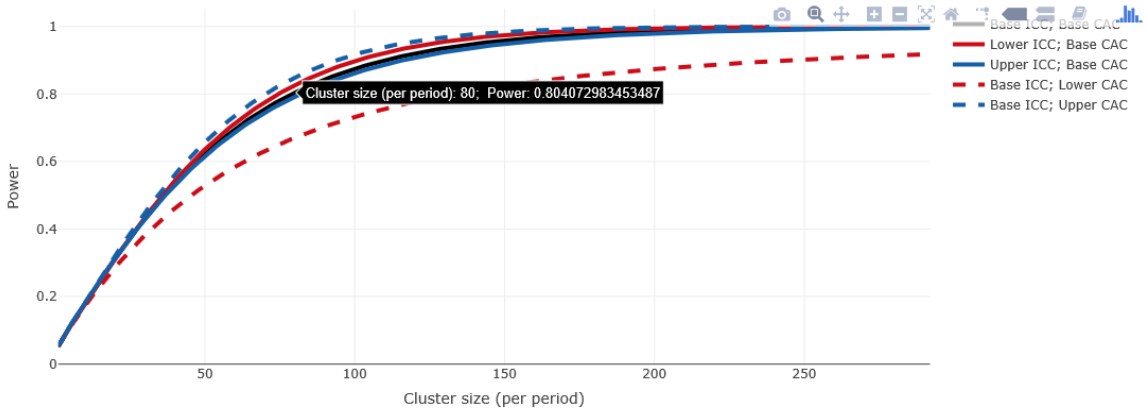
Multi cross-over

https://clusterrcts.shinyapps.io/rshinyapp/



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

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!

Cross overs, stepped wedges, and staircases! Oh my!

- Over the past few years an enhanced understanding of longitudinal cluster randomised trials has emerged
  - More realistic within-cluster correlation structures;
  - More efficient designs.
- R makes it so easy to develop tools to help researchers explore our results!
  - I try to develop an R Shiny app for each paper I write.
- There is more work to be done!
  - We can do more to get our results into the hands of researchers designing these studies.
  - How can we help researchers plan the most efficient design for their setting?

Cross overs, stepped wedges, and staircases! Oh my!

- Over the past few years an enhanced understanding of longitudinal cluster randomised trials has emerged
  - More realistic within-cluster correlation structures;
  - More efficient designs.
- R makes it so easy to develop tools to help researchers explore our results!
  - I try to develop an R Shiny app for each paper I write.
- There is more work to be done!
  - We can do more to get our results into the hands of researchers designing these studies.
  - How can we help researchers plan the most efficient design for their setting?

Questions? Ask now, or get in touch:

jessica.kasza@monash.edu

Outline

The catalyst

- In December 2017, Dr. Kristian Lum published “Statistics, we have a problem” on Medium.
  - Detailed her experience of sexual harassment at the hands of a senior colleague at various statistical conferences in the USA.
- Generated a lot of discussion on ANZstat, the Australian and New Zealand statistics mailing list.

The Statistical Society of Australia's response

- Statistical Society of Australia, SSA:
  - $\approx$ 900 members: the professional society for Australian statisticians
  - Branches in ACT, NSW, QLD, SA, VIC, WA
  - Australian statistical conferences every two years; Young Statistician conferences in off years.
- My role in the SSA: Vice-President nationally; President of the Victorian Branch 2017-2018; vice-chair of 2018 conference.
- I emailed Scott Sisson, then the President of the Statistical Society of Australia
  - What can the SSA do to prevent and respond to this sort of behaviour?
  - How can I help?

Forming a committee

- Scott agreed that the SSA needed to clearly indicate that such behaviour would not be tolerated by the society.
  - Adrian Barnett (then SSA Vice President, now President) agreed.
- I was invited to chair a committee:
  - Make recommendations about how the SSA could prevent sexual harassment at our conferences, and how the SSA could respond to any such incidents.
- The committee: Safe SSA
  - Sent out a call for members through ANZstat
  - Male and female SSA members from around Australia at various career stages.
- A lot of other statistical societies also worked hard to address this problem: ASA, ISBA

Safe SSA committee aims

1. What are the current requirements for SSA members?
2. What do we require of non-members who attend events?
3. What are other similar organisations doing?
4. Recommend policies that SSA can implement to prevent unacceptable behaviour at our events.
5. Recommend mechanisms for reporting unacceptable behaviour.
6. What else can the SSA do to provide a welcoming and safe environment free from harassment of all kinds?

Safe SSA committee aims

1. **What are the current requirements for SSA members?**
2. What do we require of non-members who attend events?
3. What are other similar organisations doing?
4. **Recommend policies that SSA can implement to prevent unacceptable behaviour at our events.**
5. **Recommend mechanisms for reporting unacceptable behaviour.**
6. What else can the SSA do to provide a welcoming and safe environment free from harassment of all kinds?

Aim 1. Requirements of SSA members

- The old SSA Code of Conduct:
  - “Members shall uphold the reputation of the profession. . . ”
  - “Members shall act with integrity towards fellow statisticians. . . and shall avoid engaging in any activity which is incompatible with their professional status”
  - No explicit mention of sexual harassment, racism, or other forms of harassment. . . . no mention made of fraudulent behaviour either!

Aim 1. Requirements of SSA members

- Updated Code of Conduct:
  - “Members shall refrain from engaging in unacceptable behaviour, including, but not limited to sexual harassment, stalking, and harassment including verbal comments relating to gender, sexual orientation, disability, race, ethnicity, religion (or lack thereof), age, national origin, gender identity or expression, or physical appearance.”
- Breaches of code may be reported to and recorded by the SSA.
  - When selecting conference speakers and Society awardees, this record will be considered.

Aim 4. Safeguarding SSA events

- All SSA-sponsored and run events must have a Code of Conduct (CoC)
  - It must explicitly list unacceptable behaviour
  - Members of a response team must be listed with contact details
  - Consequences of breaches of the Code must be listed.
- The SSA's 2018 conference (joint with the International Society for Clinical Biostatistics) was the first SSA conference to have a Code of Conduct in place.
  - I learnt a lot about conference Codes of Conduct!
  - Some breaches of the CoC were recorded.

The SSA Conference Code of Conduct

- Adherence to the CoC was a requirement for all attendees, and covered all aspects of the conference
  - Including social events and communication on social media
- Unacceptable behaviours were explicitly listed, for example:
  - Harassment, including verbal comments relating to gender, sexual orientation, disability, race, ethnicity, religion, age, national origin, gender identity or expression, or physical appearance;
  - Inappropriate and/or unwanted physical contact. . .
- Response team members listed, with photos and email addresses.

Aim 5. Reporting unacceptable behaviour

- Form a Member Conduct Committee tasked with collating reports of inappropriate behaviour
  - Make recommendations to the SSA Executive Committee for responding to reports.
- Reports may come from events and conferences, members' workplaces, or individuals
  - Support our members to come forward with reports and concerns
- A range of actions can be taken.
  - From no action to public statements about the incident.

What have I learned?

- It is important that professional societies lead the way.
- The statistical community is keen to embrace change!
  - But change takes time.
- This is just the start:
  - I want the statistical community to be safe, inclusive, and welcoming for all.
- If you see something, say something:
  - Have the uncomfortable discussions.
  - If there is something that you think needs to change in your community, ask if you can help to change it.

Please contact me if you have suggestions or concerns:

jessica.kasza@monash.edu

@JessKasza

Women in Mathematics and Statistics Conference 2021

- September 2021, Monash University.
- 3 days of talks by women working in maths, stats, data science.
- Dr Kristian Lum a keynote speaker!

`https://www.austms.org.au/WIMSIG-conference-2020`

Outline

Statistical models for Y_{kti}: subject i; period t; cluster k.

- Model 1: constant correlation.

$$Y_{kti} = \beta_t + \theta X_{kt} + C_k + \epsilon_{kti}, \quad C_k \sim N(0, \sigma_C^2), \quad \epsilon \sim N(0, \sigma_\epsilon^2)$$

- Model 2: block exchangeable.

$$Y_{kti} = \beta_t + \theta X_{kt} + C_k + CP_{kt} + \epsilon_{kti}, \quad C_k \sim N(0, \sigma_C^2), \quad CP_{kt} \sim N(0, \sigma_{CP}^2), \quad \epsilon \sim N(0, \sigma_\epsilon^2)$$

- Model 3: discrete-time decay.

$$Y_{kti} = \beta_t + \theta X_{kt} + C_{kt} + \epsilon_{kti}, \quad \epsilon \sim N(0, \sigma_\epsilon^2)$$

$$C_k = (C_{k1}, \ldots, C_{kT}) \sim N(\mathbf{0}, \sigma_C^2 \Sigma), \quad \Sigma[t, s] = r^{|t-s|}$$

Selected references

Decaying correlation structures:

- Grantham, Kasza, Heritier, Hemming, Forbes. *Accounting for a decaying correlation structure in cluster randomized trials with continuous recruitment*. **Statistics in Medicine**. 2019.
- Kasza, Hemming, Hooper, Matthews, Forbes. *Impact of non-uniform correlation structure on sample size and power in multiple-period cluster randomised trials*. **Statistical Methods in Medical Research**. 2019.

Impact of misspecification:

- Kasza, Forbes. *Inference for the treatment effect in multiple-period cluster randomised trials when random effect correlation structure is misspecified*. **Statistical Methods in Medical Research**. 2019.

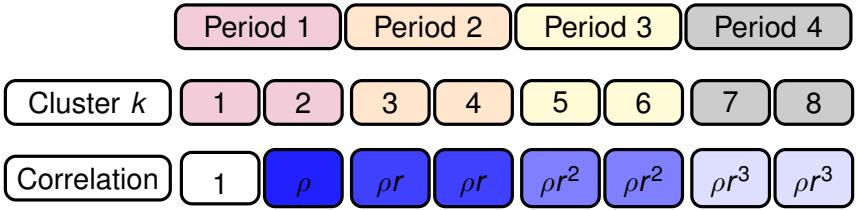
Information content of stepped wedge designs:

- Kasza, Forbes. *Information content of cluster-period cells in stepped wedge trials*. **Biometrics**. 2019.
- Kasza, Taljaard, Forbes. *Information content of stepped-wedge designs when treatment effect heterogeneity and/or implementation periods are present*. **Statistics in Medicine**. 2019.

Crossing over:

- Grantham, Kasza, Heritier, Hemming, Litton, Forbes. *How many times should a cluster randomised trial cross over?* **Statistics in Medicine**. 2019.

Decaying between-period ICCs



- The correlation between participant outcomes decreases the further their measurement periods are apart in time.
  - For participants in period t and s: correlation is $\rho r^{|t-s|}$

NOTE: Shading indicates the degree of correlation between participant 1 and other participants. The correlation of participant 1 with themselves is 1.

Statistical Models

Y_{kti}: subject i; period t; cluster k.

- Constant correlation model:

$$Y_{kti} = \beta_t + \theta X_{kt} + C_k + \epsilon_{kti}, \quad C_k \sim N(0, \sigma_C^2), \quad \epsilon \sim N(0, \sigma_\epsilon^2)$$

- Block exchangeable model:

$$Y_{kti} = \beta_t + \theta X_{kt} + C_k + CP_{kt} + \epsilon_{kti}, \quad C_k \sim N(0, \sigma_C^2), \quad CP_{kt} \sim N(0, \sigma_{CP}^2), \quad \epsilon \sim N(0, \sigma_\epsilon^2)$$

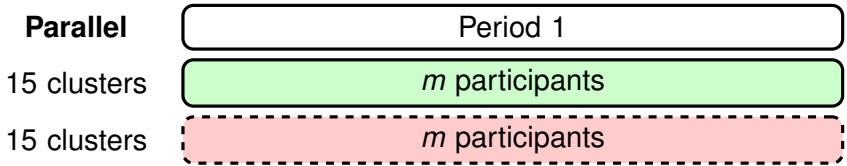
- Discrete-time decay model:

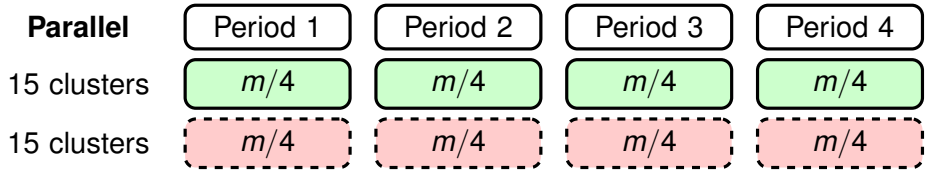
$$Y_{kti} = \beta_t + \theta X_{kt} + C_{kt} + \epsilon_{kti}, \quad \epsilon \sim N(0, \sigma_\epsilon^2)$$

$$C_k = (C_{k1}, \ldots, C_{kT}) \sim N(\mathbf{0}, \sigma_C^2 \Sigma), \quad \Sigma[t, s] = r^{|t-s|}$$

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.