Inspired by research. Driven by compassion. Inspiré par la recherche. Guidé par la compassion.

STEPPED WEDGE CLUSTER RANDOMIZED TRIALS: WHAT, HOW AND WHEN?

NIA IMPACT COLLABORATORY GRAND ROUNDS
19 December 2019

NIA IMPACT COLLABORATORY
TRANSFORMING DEMENTIA CARE

MONICA TALJAARD

Senior Scientist, Ottawa Hospital Research Institute
Associate Professor, University of Ottawa

The Ottawa Hospital | L'Hôpital d'Ottawa

OUTLINE

1. Refresher: Cluster randomized trials (CRTs)
2. What is a stepped wedge cluster randomized trial (SW-CRT)?
3. Analysis of SW-CRTs
4. Sample size calculation for SW-CRTs
5. What is an appropriate justification for using a SW-CRT?
6. Summary

NIA IMPACT COLLABORATORY
TRANSFORMING DEMENTIA CARE

CLUSTER RANDOMIZED TRIALS

- What is a cluster randomized trial (CRT)?
  - Units of randomization are intact groups (“clusters”) rather than individuals
  - Outcomes are observed on multiple individuals within each cluster
- Key characteristics:
  - Multiple observations from the same cluster usually positively correlated
  - The strength of the correlation can be measured by the Intracluster Correlation Coefficient (ICC)
  - Must account for ICC in both sample size calculation and analysis to obtain valid inferences

A DEFINITION OF ICC

- Assume the outcome *Y* is continuous with variance σ^2
- The variance σ^2 may be expressed as the sum of two components:

$$\sigma^2 = \sigma_b^2 + \sigma_w^2$$

where

σ^2_b = variance between cluster means

σ^2_w = variance of individuals within clusters

- Then the ICC is defined as

$$\rho = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_w^2}; \qquad 0 \le \rho \le 1$$

QUANTIFYING THE EFFECTS OF CLUSTERING

- In a standard clinical trial with *n* individuals randomized to each arm, we have:

$$Var\left(\bar{Y}_i\right) = \frac{\sigma^2}{n}, \qquad i = 1,2$$

- In a CRT with $n=km$ individuals per arm (where k = number of clusters, and m=number of individuals per cluster), we have:

$$Var\left(\bar{Y}_i\right) = \frac{\sigma^2}{km}\left[1+\left(m-1\right)\rho\right]$$

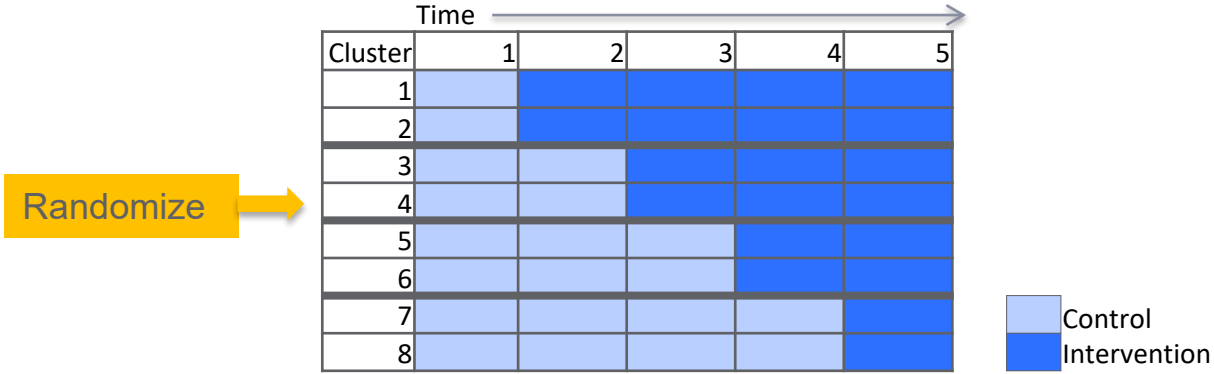
- The variance inflation factor $1+(m-1)\rho$ is called the "Design Effect"
- Sample size for a CRT may be obtained my multiplying *n* under individual randomization by the Design Effect (+ any necessary small sample correction)

WHAT IS A STEPPED WEDGE CRT (SW-CRT)?

- A novel type of CRT design – often used to evaluate health system and service delivery interventions
- Rapid rise in popularity
- Methods not fully developed
- Quality of published trials has been poor

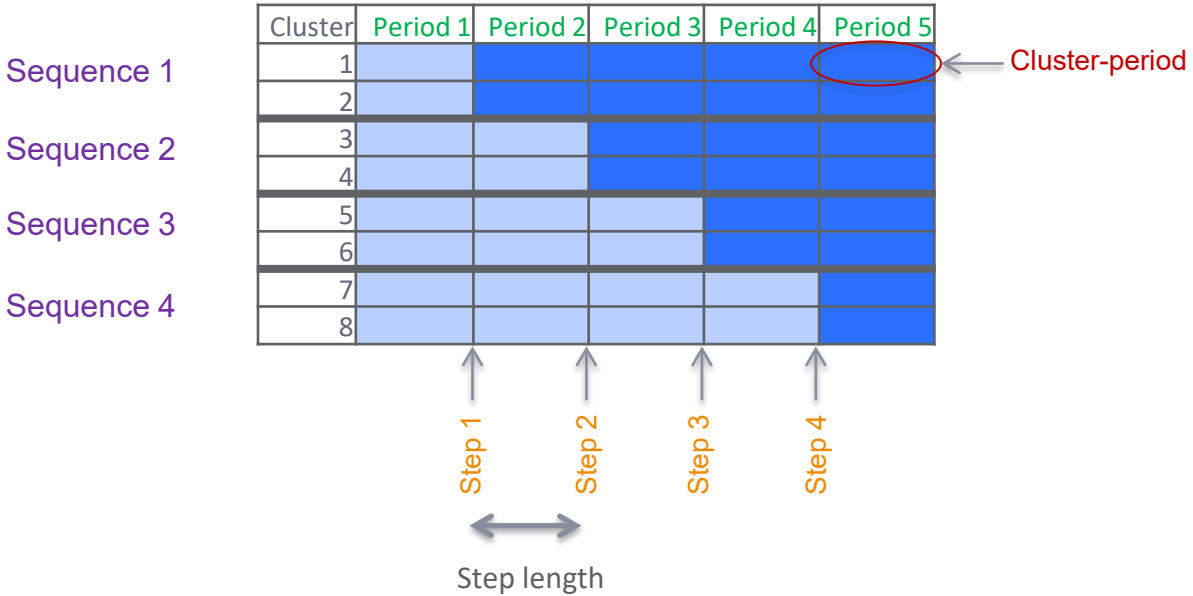
- Martin J, Taljaard M, Girling A, et al. Systematic review finds major deficiencies in sample size methodology and reporting for stepped-wedge cluster randomised trials. *BMJ Open* 2016;6:e010166
- Grayling MJ, Wason JM, Mander AP. Stepped wedge cluster randomized controlled trial designs: a review of reporting quality and design features. *Trials* 2017;18:33.

THE STANDARD SW-CRT DESIGN



- Sequential and unidirectional cross-over of clusters (or groups of clusters)
- Clusters are randomized to different (calendar) times of crossing over
- Outcomes are assessed repeatedly in each cluster

TERMINOLOGY

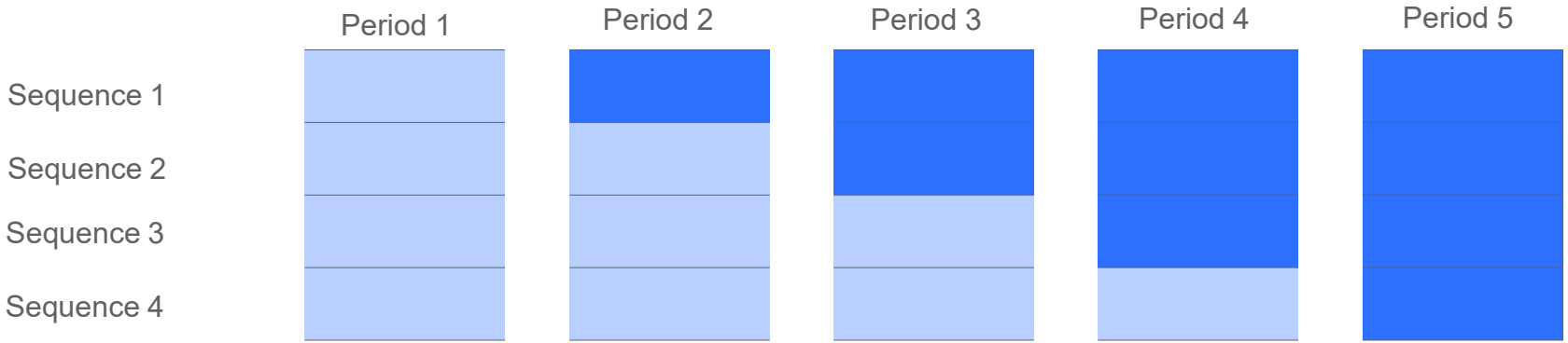


THREE MAIN TYPES OF SW-CRT DESIGNS

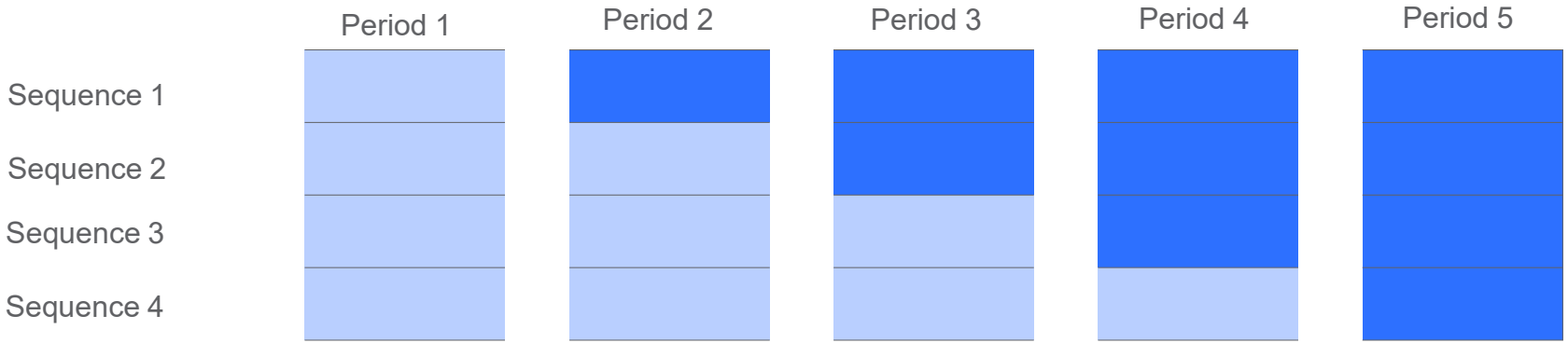
- Three main types of SW-CRT designs
  1. Closed cohort design
  2. Continuous recruitment short exposure design
  3. Repeated cross-section or open cohort design

Copas AJ e.a. (2015) Designing a stepped wedge trial: three main designs, carry-over effects and randomisation approaches. *Trials*; 16:352

1) CLOSED COHORT DESIGN



1) CLOSED COHORT DESIGN

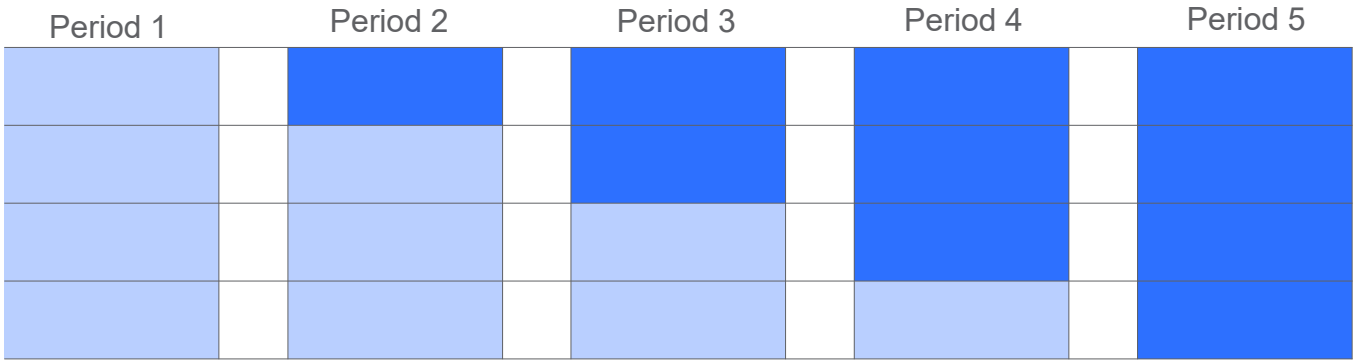


Timeline

Cluster Recruitment

NIA IMPACT COLLABORATORY
TRANSFORMING DEMENTIA CARE

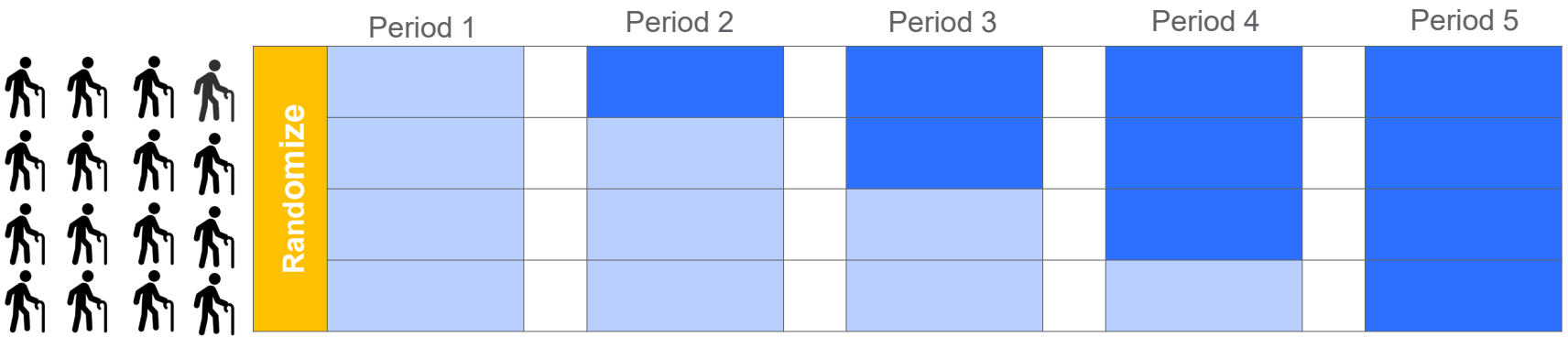
1) CLOSED COHORT DESIGN



Timeline

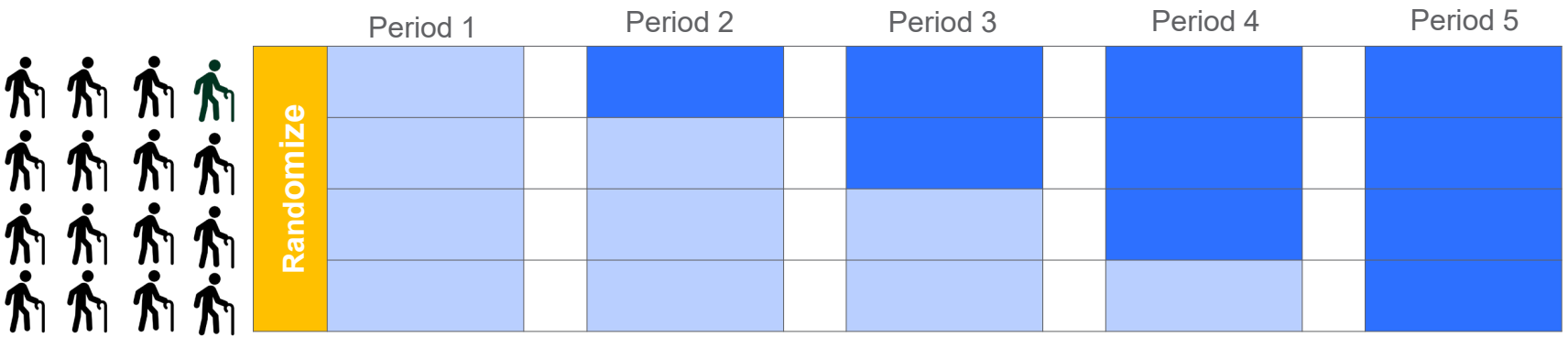
Individual Recruitment

1) CLOSED COHORT DESIGN



Timeline

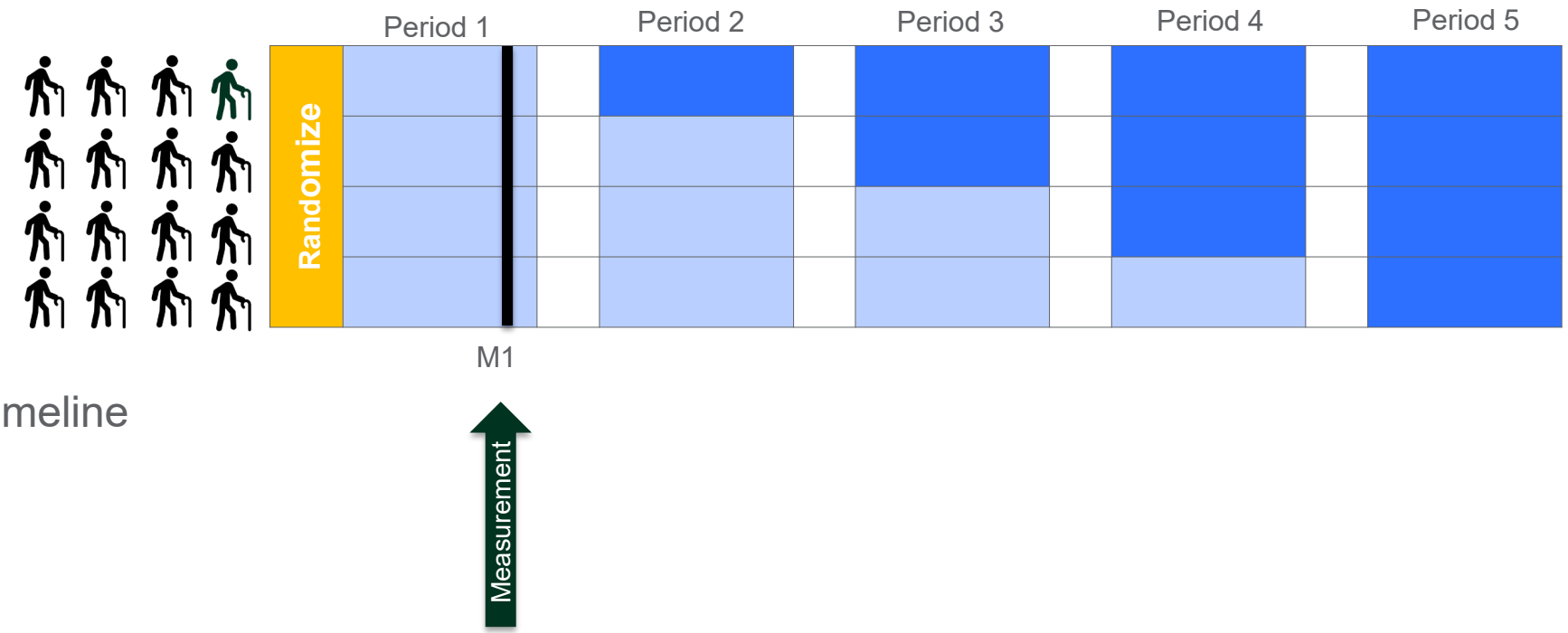
1) CLOSED COHORT DESIGN



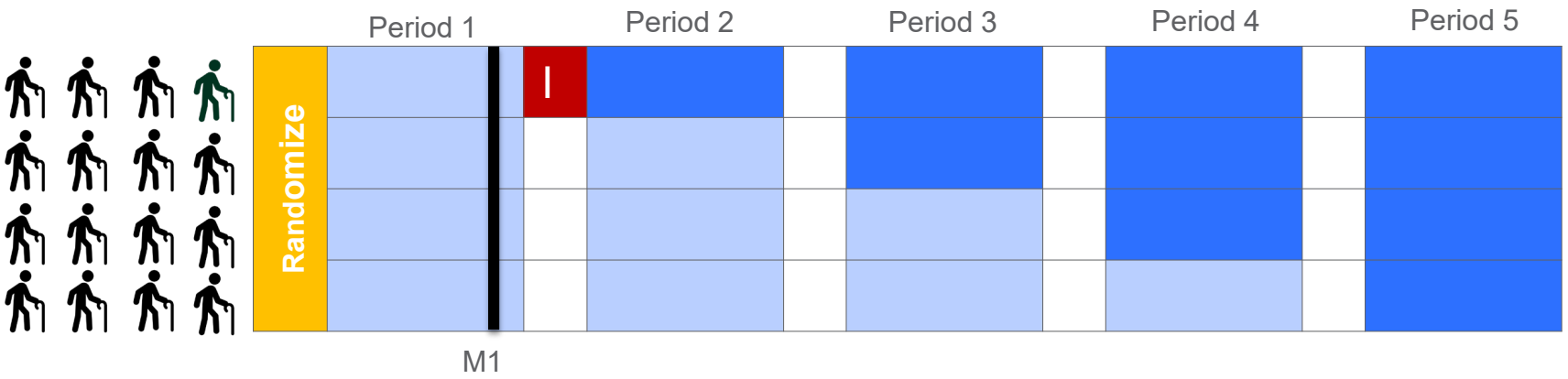
Timeline

Exposed to control

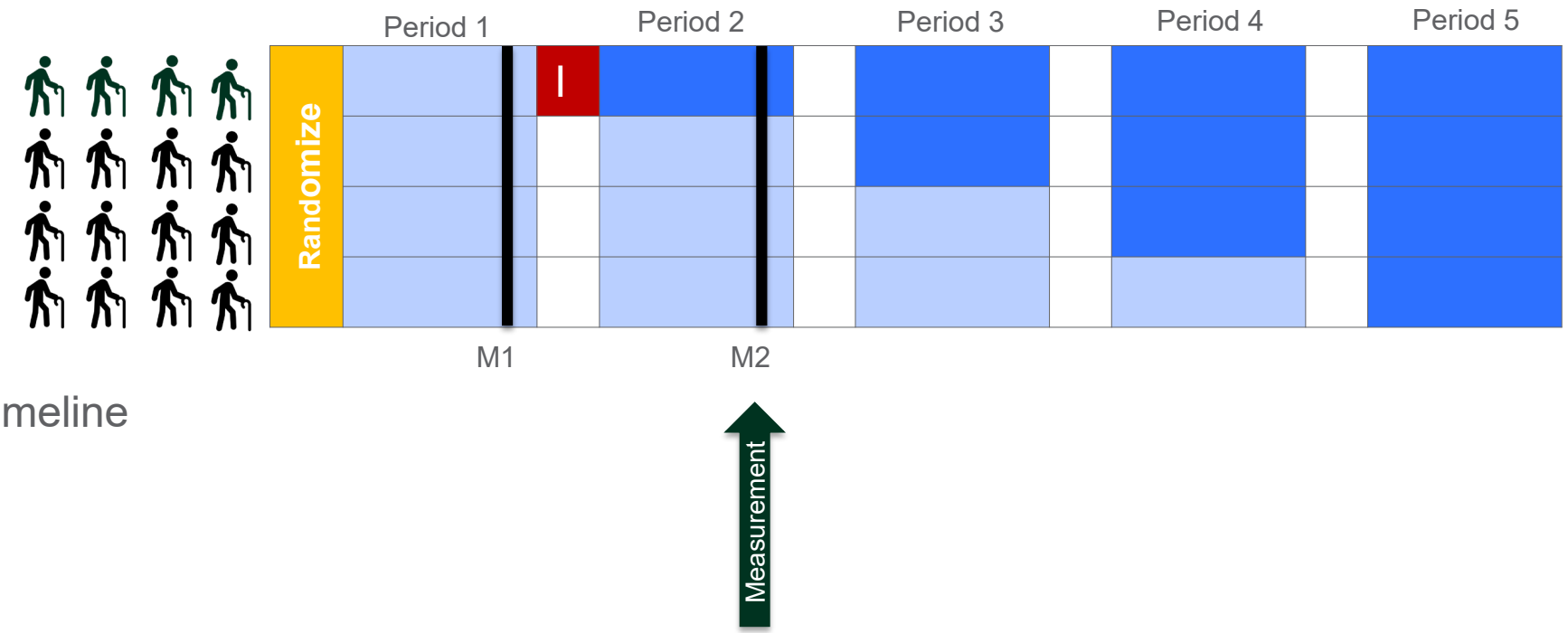
1) CLOSED COHORT DESIGN



1) CLOSED COHORT DESIGN

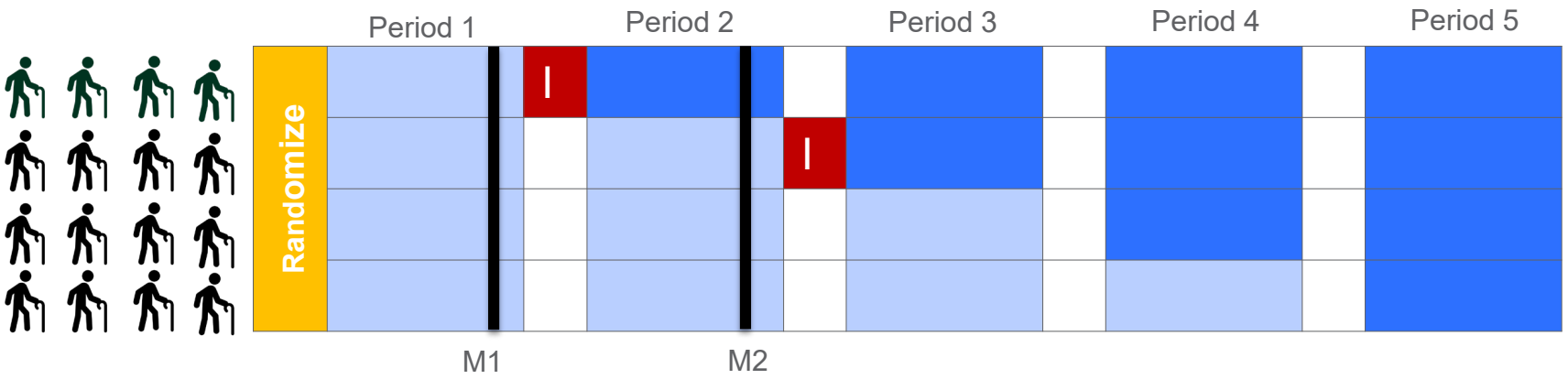


1) CLOSED COHORT DESIGN



Timeline

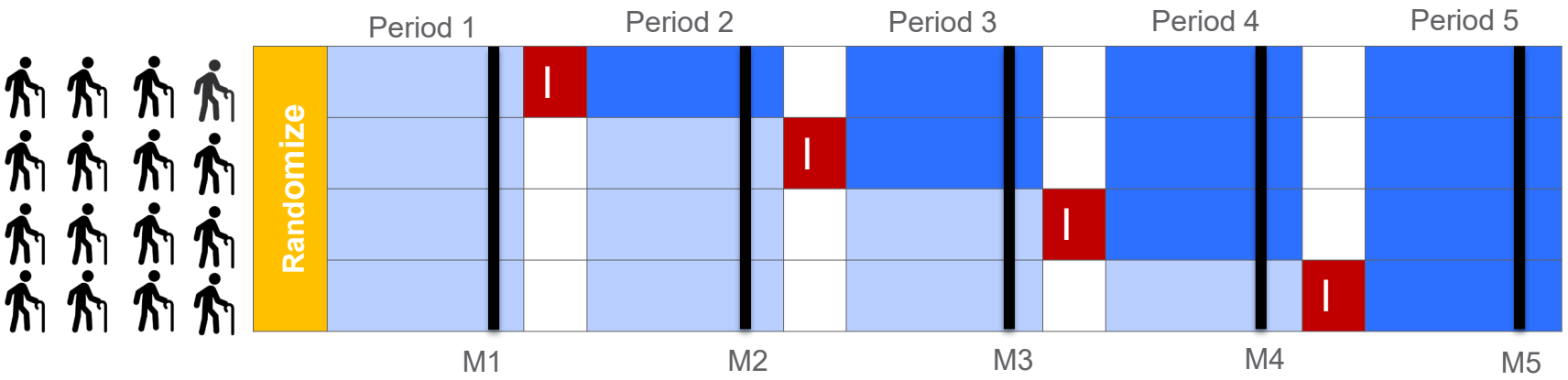
1) CLOSED COHORT DESIGN



Timeline

Intervention delivery

1) CLOSED COHORT DESIGN



- Note: In the most basic version of the design, we have to assume…
  - Once intervention has been delivered, it keeps working! (no decay effects)
  - Intervention works immediately! (no learning or lagged effects)

1) CLOSED COHORT DESIGN

- Summary
  - Participants are recruited at the beginning of the trial and participate to the end
  - Each participant is exposed to both control and intervention conditions
  - The same participant is measured repeatedly throughout the trial

EXAMPLE 1: CLOSED COHORT

European Journal of Internal Medicine 28 (2016) 43–51

Contents lists available at ScienceDirect

European Journal of Internal Medicine

journal homepage: www.elsevier.com/locate/ejim

Original Article

Effectiveness of a Geriatric Care Model for frail older adults in primary care: Results from a stepped wedge cluster randomized trial

Emiel O. Hoogendijk [a,b,c,*], Henriëtte E. van der Horst [a], Peter M. van de Ven [c], Jos W.R. Twisk [c], Dorly J.H. Deeg [c], Dinnus H.M. Frijters [a], Karen M. van Leeuwen [a,d], Jos P.C.M. van Campen [e], Giel Nijpels [a], Aaltje P.D. Jansen [a], Hein P.J. van Hout [a]

[a] Department of General Practice & Elderly Care Medicine, EMGO + Institute for Health and Care Research, VU University Medical Center, Amsterdam, The Netherlands
[b] Gérontopôle, Department of Internal Medicine and Geriatrics, Toulouse University Hospital, Toulouse, France
[c] Department of Epidemiology & Biostatistics, EMGO + Institute for Health and Care Research, VU University Medical Center, Amsterdam, The Netherlands
[d] Department of Health Sciences, EMGO+ Institute for Health and Care Research, Faculty of Earth & Life Sciences, VU University, Amsterdam, The Netherlands
[e] Department of Geriatric Medicine, Slotervaart Hospital, Amsterdam, The Netherlands

EXAMPLE 1: CLOSED COHORT

- **Objective**: Evaluate a multifaceted geriatric primary care model for community-dwelling frail older adults
- **Design:** SW-CRT in 35 primary care practices in the Netherlands over 24 months (1,147 patients)
- **Intervention**: Geriatric in-home assessment and visits by a practice nurse plus a tailored care plan overseen by a geriatric expert team
- **Control**: Usual care
- **Primary outcome**: Quality of Life assessed on the same individuals every six months using computer assisted personal interviewing
- **Results**: No beneficial effects

EXAMPLE 1: CLOSED COHORT

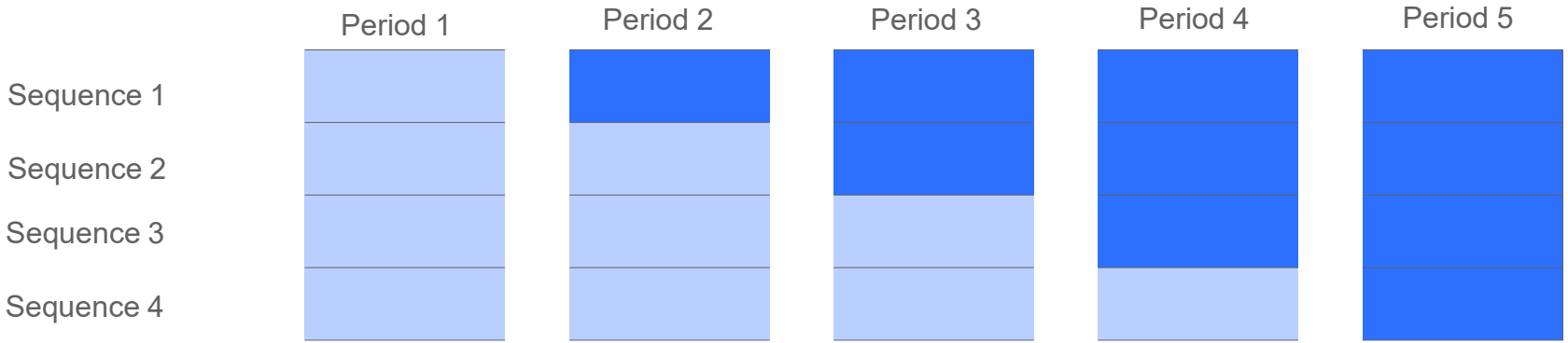
Clusters	Group	Follow-up time				
		Baseline	6 months	12 months	18 months	24 months
10	Group 1	Control	6 months	12 months	18 months	24 months
9	Group 2	Control	Control	6 months	12 months	18 months
8	Group 3	Control	Control	Control	6 months	12 months
8	Group 4	Control	Control	Control	Control	6 months

Control = control measurements

6 months = time since the start of the intervention

- Comments:
  - "Practices were randomized… before patient recruitment started"
  - "One practice in allocation group 4 did not start the intervention"
  - "31.8% of patients did not complete the 24-month study"

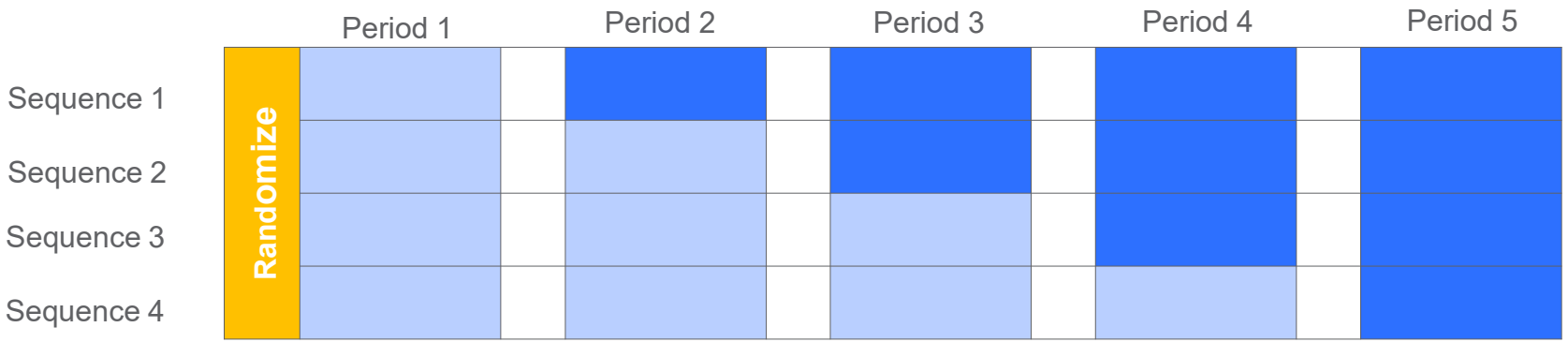
2) CONTINUOUS RECRUITMENT SHORT EXPOSURE



Timeline

Cluster Recruitment

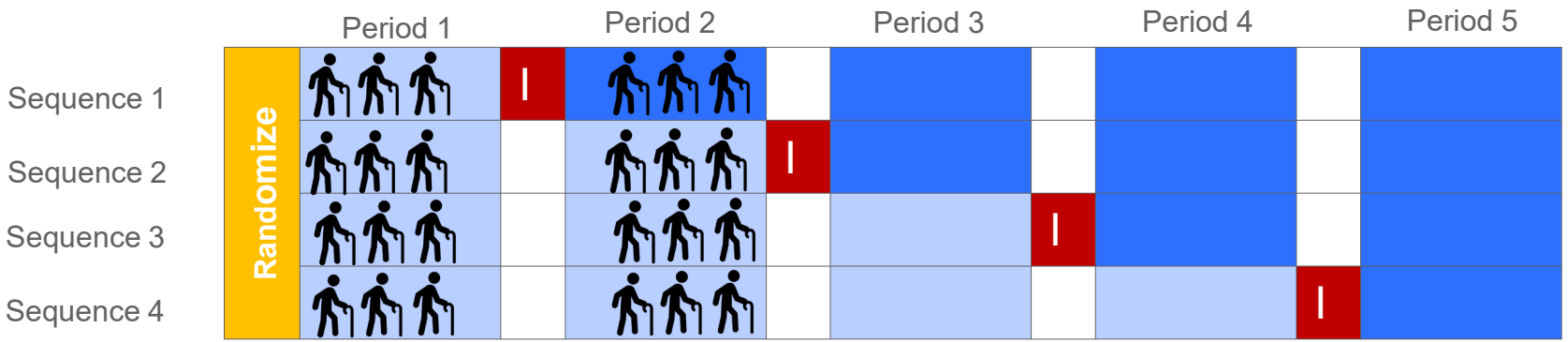
2) CONTINUOUS RECRUITMENT SHORT EXPOSURE



Timeline

Cluster Recruitment

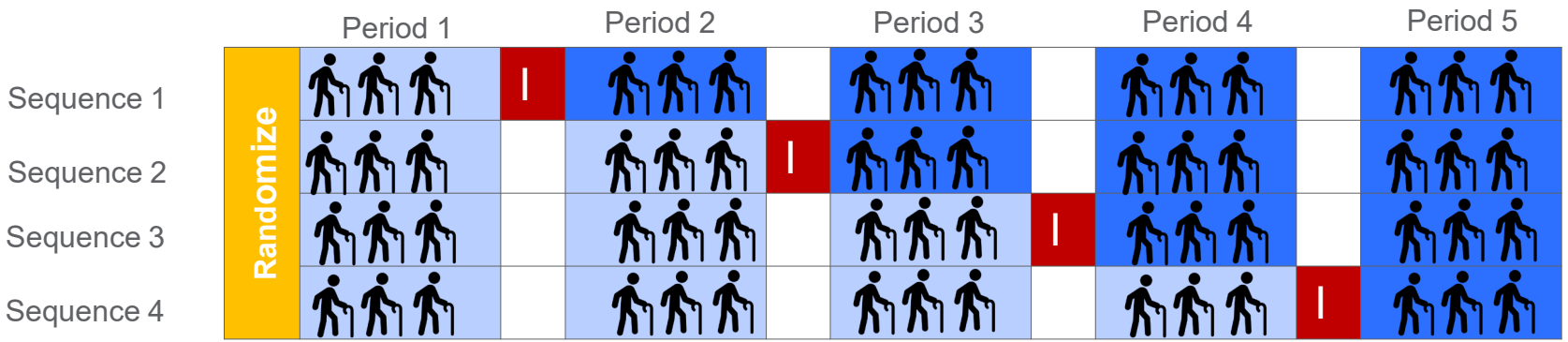
2) CONTINUOUS RECRUITMENT SHORT EXPOSURE



Timeline

Individual recruitment

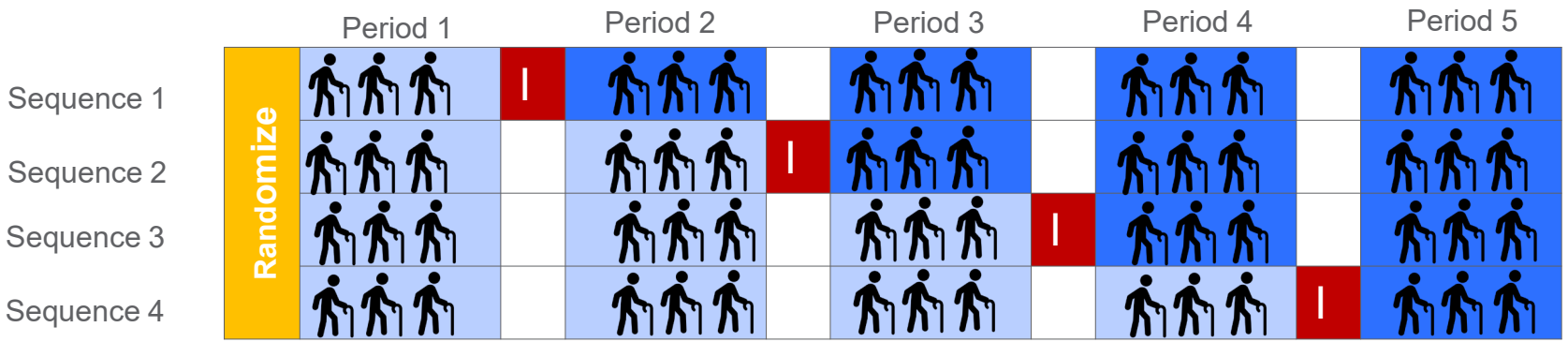
2) CONTINUOUS RECRUITMENT SHORT EXPOSURE



Timeline

Individual recruitment

2) CONTINUOUS RECRUITMENT SHORT EXPOSURE



- Note: Risk of within-cluster contamination increases when…
  - Duration of exposure is long
  - There is no allowance for a transition period

2) CONTINUOUS RECRUITMENT SHORT EXPOSURE

- Summary
  - Participants are identified and become exposed on a continuous basis
  - Each participant exposed to either control or intervention – not both
  - Different participants measured in each cluster over time

EXAMPLE 2: CONT RECRUITMENT SHORT EXPOSURE

Open access Protocol

BMJ Open

Stepped wedge cluster randomised controlled trial to assess the effectiveness of an optimisation strategy for general anaesthesia on postoperative morbidity and mortality in elderly patients (the OPTI-AGED study): a study protocol

Serge Molliex,[1] Sylvie Passot,[1] Emmanuel Futier,[2] Marlène Bonnefoi,[3] Florence Rancon,[3] Yannick Lemanach,[4] Bruno Pereira[5]

Molliex S, *et al*. *BMJ Open* 2018;**8**:e021053. doi:10.1136/bmjopen-2017-021053

EXAMPLE 2: CONT RECRUITMENT SHORT EXPOSURE

- **Objective**: Evaluate a multifaceted general anaesthesia optimisation strategy in elderly patients undergoing high-risk surgery
- **Design**: SW-CRT in 27 French hospitals over 24 months (2,500 patients)
- **Intervention**: Optimisation of general anaesthesia (haemodynamic intervention, lung-protective ventilation and electroencephalographic monitoring of anaesthesia depth)
- **Control**: Usual care
- **Primary outcome**: Composite of major post-operative complications or mortality on day of surgery, day 7, day 30, and 1 year post-surgery

EXAMPLE 2: CONT RECRUITMENT SHORT EXPOSURE

Table 1 Stepped wedge study design of OPTI-AGED trial

Clusters	Initiation	Step1	Step2	Step3	Step4	Step5
5-6	Control	Control	Control	Control	Control	Intervention
5-6	Control	Control	Control	Control	Intervention	Intervention
5-6	Control	Control	Control	Intervention	Intervention	Intervention
5-6	Control	Control	Intervention	Intervention	Intervention	Intervention
5-6	Control	Intervention	Intervention	Intervention	Intervention	Intervention

Six intervals of 4 months will be fixed over 24 months.
The randomisation will involve five steps for which 5–6 centres will be included in each cluster.

- Comments:
  - "…training on the intervention will be performed in each center within 15 days preceding the cross-over…"
  - Rationale for choosing a SW-CRT: "It is unethical to withhold an intervention anticipated to be beneficial"

3) OPEN COHORT

- Many individuals exposed from the start; some may leave and others may become eligible over time
- Variation 1:
  - Measurements are taken on a small fraction of individuals within large clusters at discrete calendar times (unlikely that any one individual is measured more than once)
- Variation 2:
  - Measurements taken repeatedly on all eligible individuals in every period (likely that many or at least some individuals are measured multiple times under both control and intervention conditions)

EXAMPLE 3: OPEN COHORT

Stern *et al. BMC Health Services Research* 2014, **14**:83
http://www.biomedcentral.com/1472-6963/14/83

RESEARCH ARTICLE Open Access

Pressure ulcer multidisciplinary teams via telemedicine: a pragmatic cluster randomized stepped wedge trial in long term care

Anita Stern[1*], Nicholas Mitsakakis[2], Mike Paulden[3], Shabbir Alibhai[4], Josephine Wong[2], George Tomlinson[5], Ann-Sylvia Brooker[2], Murray Krahn[2] and Merrick Zwarenstein[6]

EXAMPLE 3: OPEN COHORT

- **Objective**: Evaluate the effectiveness of enhanced multidisciplinary teams for the treatment of pressure ulcers in long term care facilities in Ontario, Canada
- **Design**: SW-CRT in 12 facilities (137 residents with 259 pressure ulcers) over 17 months
- **Intervention**: Visit by advance practice nurse; staff education; support by an off-site hospital based expert multi-disciplinary wound care team via email, telephone, or video link
- **Control**: Usual care
- **Primary outcome:** Pressure ulcer surface area measured by a blinded assessor who visited facilities every 2 weeks to take photographs
- **Results**: No statistically significant difference

EXAMPLE 3: OPEN COHORT

Table 1 Study design

Time (Months)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Facility number 12	-	-	-	-	-	-	-	c	c	c	c	c	c	P1	P1	P1	P2
11	-	-	-	-	-	-	c	c	c	c	c	P1	P1	P1	P2	P2	P2
10	c	c	c	c	c	c	c	c	c	c	c	c	P1	P1	P1	P2	P2
9	c	c	c	c	c	c	c	c	c	c	c	P1	P1	P1	P2	P2	P2
8	c	c	c	c	c	c	c	c	c	c	P1	P1	P1	P2	P2	P2	P2
7	c	c	c	c	c	c	c	c	c	P1	P1	P1	P2	P2	P2	P2	P2
6	c	c	c	c	c	c	c	c	P1	P1	P1	P2	P2	P2	P2	P2	P2
5	c	c	c	c	c	c	c	P1	P1	P1	P2	P2	P2	P2	P2	P2	P2
4	c	c	c	c	c	c	P1	P1	P1	P2	P2	P2	P2	P2	P2	P2	P2
3	c	c	c	c	c	P1	P1	P1	P2	P2	P2	P2	P2	P2	P2	P2	P2
2	c	c	c	c	P1	P1	P1	P2	P2	P2	P2	P2	P2	P2	P2	P2	P2
1	c	c	c	P1	P1	P1	P2	P2	P2	P2	P2	P2	P2	P2	P2	P2	P2

c = Control; **P1** = Intervention- Phase 1 (onsite one day/week); **P2** = Intervention Phase 2 (primarily remote bi-weekly).

"Prevalence rates were lower than anticipated, and so 2 additional eligible facilities were randomly selected from the eligible sites and randomized."

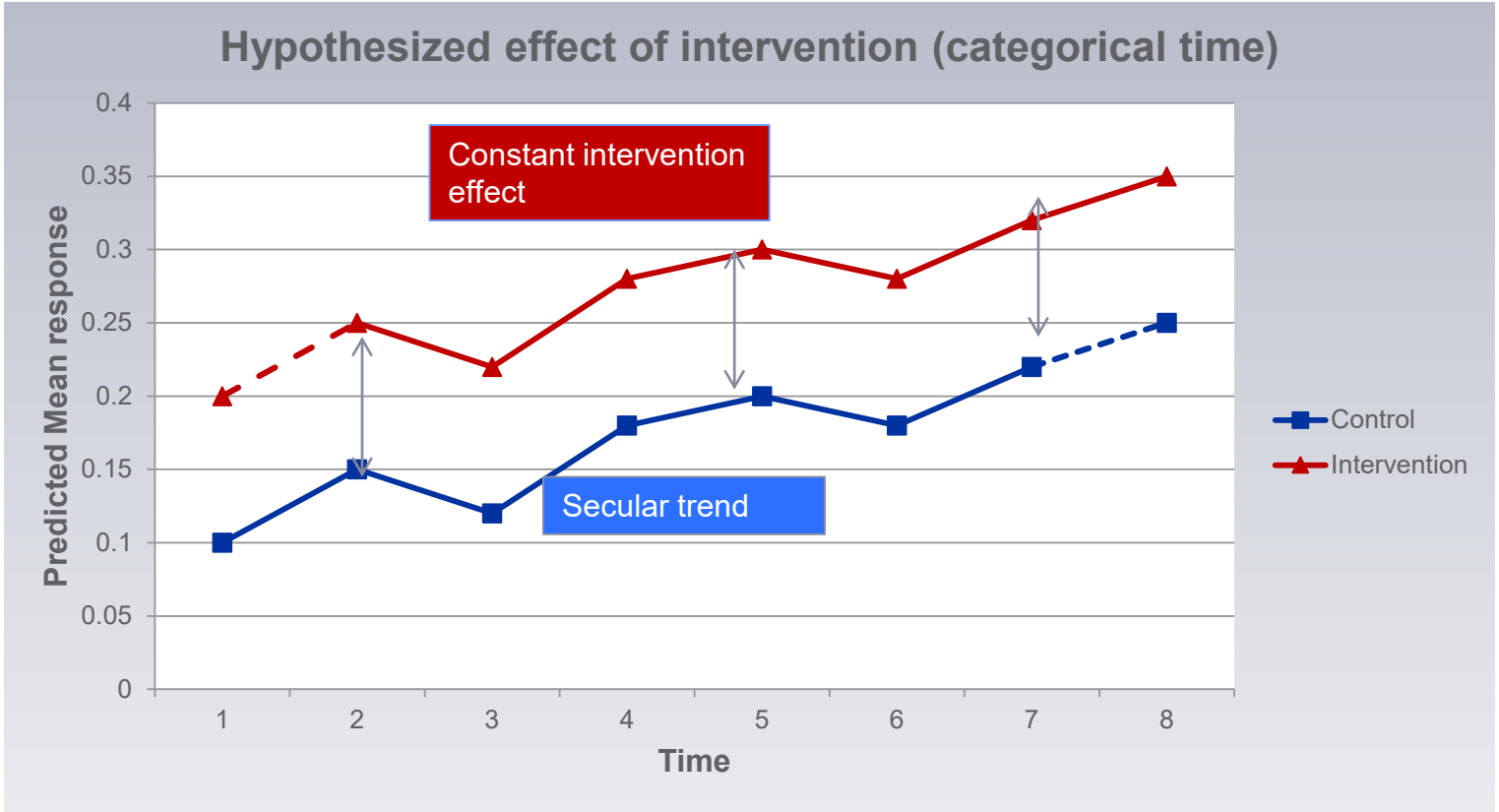
ANALYSIS OF THE SW-CRT

- Focusing here on General(ized) Linear Mixed Model (GLMM)

Essential for obtaining unbiased treatment effect

- Cross-sectional design:
  - Fixed (categorical) effect for time
  - Fixed indicator for treatment or control
  - Random intercept for cluster
  - Random time effect for cluster
- Cohort design:
  - Add random intercept for individual

IMPLICATIONS OF ASSUMED FIXED EFFECTS



IMPLICATIONS OF ASSUMED RANDOM EFFECTS

- ICC in a standard CRT (measurements taken same time)

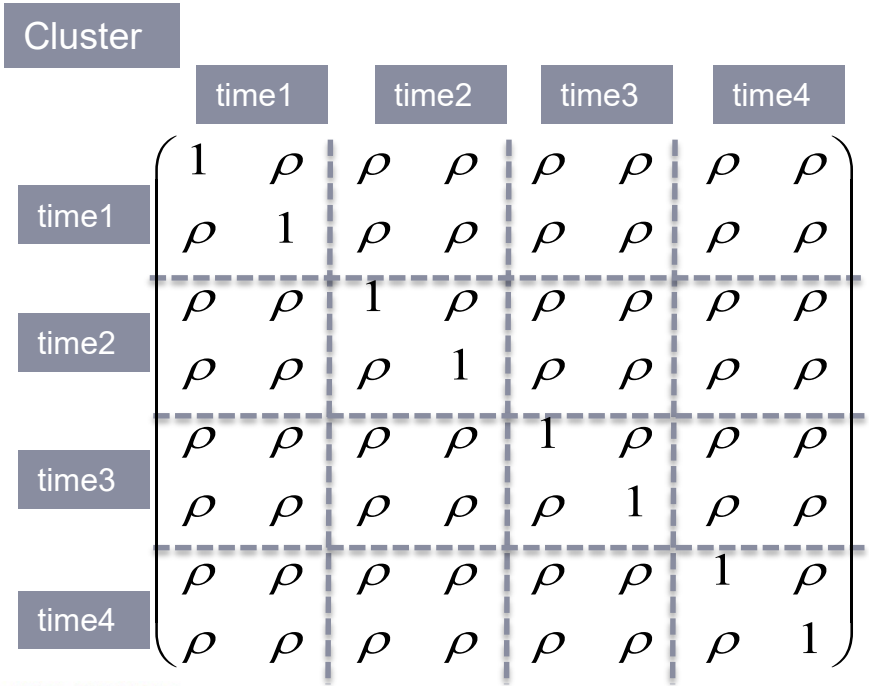
Cluster

$$\begin{pmatrix} 1 & \rho & \rho & \rho & \rho & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho & \rho & \rho & \rho & \rho \\ \rho & \rho & 1 & \rho & \rho & \rho & \rho & \rho \\ \rho & \rho & \rho & 1 & \rho & \rho & \rho & \rho \\ \rho & \rho & \rho & \rho & 1 & \rho & \rho & \rho \\ \rho & \rho & \rho & \rho & \rho & 1 & \rho & \rho \\ \rho & \rho & \rho & \rho & \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & \rho & \rho & \rho & \rho & 1 \end{pmatrix}$$

$$\rho = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_w^2}; \qquad 0 \le \rho \le 1$$

IMPLICATIONS OF ASSUMED RANDOM EFFECTS

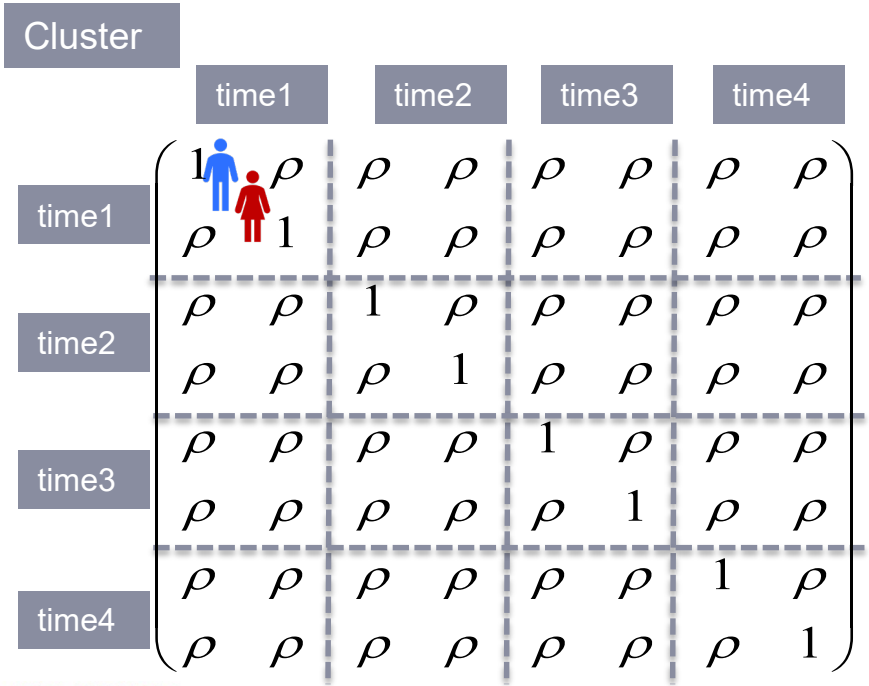
- ICC in a SW-CRT (measurements taken at 4 cross-sections in time)



- It would not make sense to assume ICC is the same, no matter how far apart measurements are

IMPLICATIONS OF ASSUMED RANDOM EFFECTS

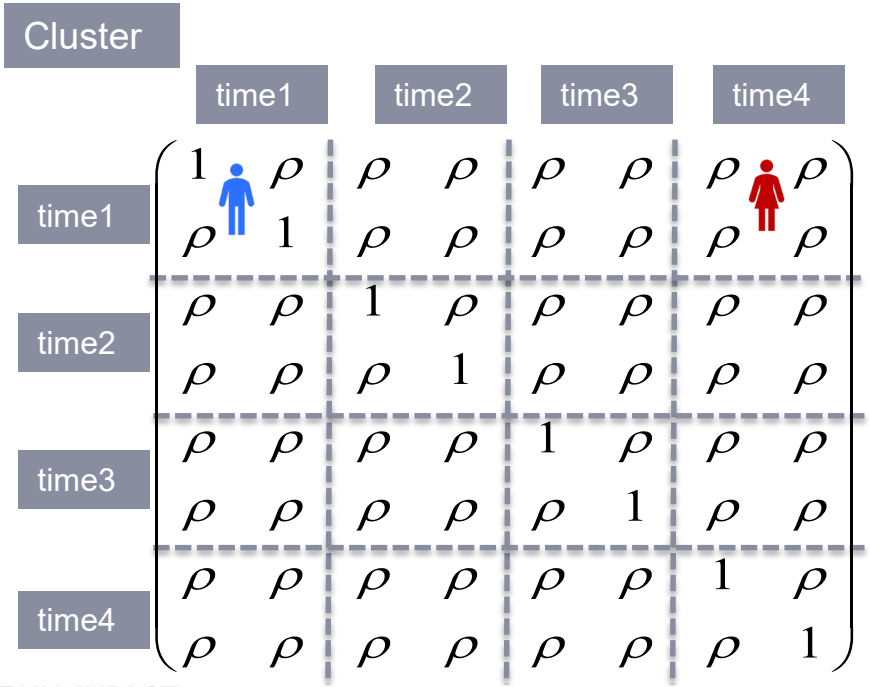
- ICC in a SW-CRT (measurements taken at 4 cross-sections in time)



- It would not make sense to assume ICC is the same, no matter how far apart measurements are

IMPLICATIONS OF ASSUMED RANDOM EFFECTS

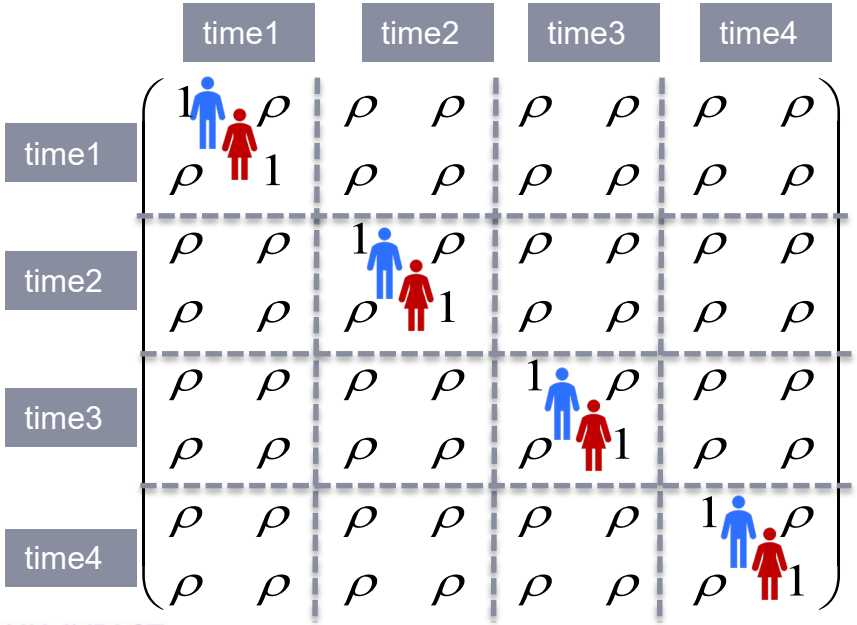
- ICC in a SW-CRT (measurements taken at 4 cross-sections in time)



- It would not make sense to assume ICC is the same, no matter how far apart measurements are

IMPLICATIONS OF ASSUMED RANDOM EFFECTS

- Define two different ICCs: within-period ICC and between-period ICC

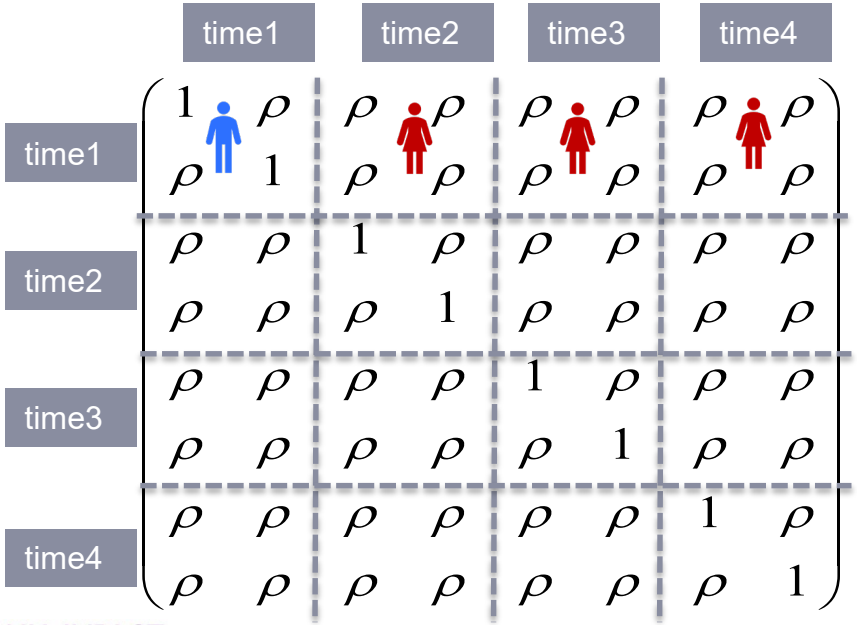


- Within-period ICC: correlation between any two individuals in the same cluster and same time

$$wpICC = \frac{\sigma_b^2 + \sigma_t^2}{\sigma_b^2 + \sigma_t^2 + \sigma_w^2} = \rho_0$$

IMPLICATIONS OF ASSUMED RANDOM EFFECTS

- Define two different ICCs: within-period ICC and between-period ICC



- Between-period ICC: correlation between any two individuals in the same cluster but different times

$$bpICC = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_t^2 + \sigma_w^2} = \omega\rho_0,$$

$$0 \le \omega \le 1$$

THE CLUSTER AUTOCORRELATION COEFFICIENT

- The ratio of the between-period and within-period ICCs is called the “Cluster Autocorrelation Coefficient” (CAC), denoted ω
  - CAC measures the extent of the correlation decay (e.g., CAC=0.8 implies a 20% decay in the correlation)
- Incorrectly assuming CAC=1 will underestimate the required sample size
- Note that earliest sample size methodology for SW-CRT did not account for the CAC

- Kasza J & Forbes A. Estimating variance components in multiple-period cluster randomised trials when random effect correlation structure is misspecified. *Stat Methods Med Res* 2018.
- Kasza J, Hemming K, Hooper R, Matthews JNS, Forbes AB et al. Impact of non-uniform correlation structure on sample size and power in multiple-period cluster randomised trials. *Stat Methods Med Res* 2017

SAMPLE SIZE CALCULATION FOR THE SW-CRT

- Here, illustrating the simplest approach using design effects
- Based on the GLMM described previously
- Works for cohort and cross-sectional designs, continuous or binary outcomes
- Refinements are possible based on more complex correlation structures
- Methodology assumes large number of clusters

- Hooper R, Bourke L. Cluster randomised trials with repeated cross sections: alternatives to parallel group designs. *BMJ*. 2015 Jun 8;350:h2925
- Hooper R et al. (2016) Sample size calculation for stepped wedge and other longitudinal cluster randomised trials. *Statistics in Medicine* 35(26):4718-4728

CALCULATION OF THE REQUIRED NUMBER OF CLUSTERS

- Five steps:
  1. Calculate **total** required sample size assuming individual randomization N_{ind}
  2. Multiply by design effect due to clustering $Deff_c = 1+(m\text{ -}1)\rho_0$
  3. Multiply by design effect due to repeated assessment $Deff_t$ (see next slide)
  4. Divide by cluster size per period (m) to determine **total** required number of clusters (k)

$$k = \frac{N_{ind} \times Deff_c \times Deff_t}{m}$$

  5. Round up to multiple of number of sequences

DESIGN EFFECT DUE TO REPEATED ASSESSMENT

- Function of number of sequences *t* and the correlation between cluster means at two different times *R* :

$$Deff_t = \frac{3t(1-R)(1+tR)}{(t^2-1)(2+tR)}$$

- *R* is defined, for cross-sectional and cohort designs, respectively as:

$$R = \frac{m\rho_0\omega}{1+(m-1)\rho_0} \qquad R = \frac{m\rho_0\omega+(1-\rho_0)\tau}{1+(m-1)\rho_0}$$

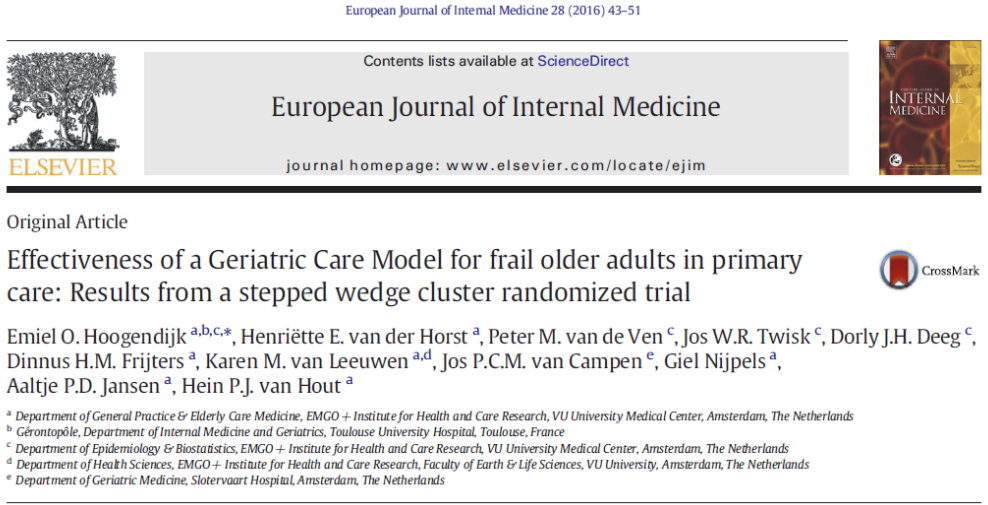
where ω is the Cluster Autocorrelation Coefficient (CAC) and τ is the Individual Autocorrelation Coefficient (IAC)

EMPIRICAL ESTIMATES FOR DESIGN PARAMETERS

- Can be challenging to obtain reliable empirical estimates for the within-period ICC and CAC
  - Ideally, fit the GLMM to raw longitudinal data with the “correct” period length (e.g., historical routinely collected data) and use the estimated variance components to calculate wpICC and CAC
  - For binary data, require estimates on the proportions (not logistic) scale
  - If no prior information, consider assuming CAC between 0.6 to 0.8
  - Essential to examine sensitivity across a range of plausible values

WORKED EXAMPLE 1: CLOSED COHORT

- Sample size parameters (as stated in manuscript):
  - 90% power, $\alpha = 0.05$
  - $t = 4$ sequences
  - m=? individuals per practice
  - Standard deviation = 7.1
  - Target difference = 3
  - "ICC" = 0.02, CAC=?, IAC=0.66
  - 20% attrition



European Journal of Internal Medicine 28 (2016) 43–51

Contents lists available at ScienceDirect

European Journal of Internal Medicine

journal homepage: www.elsevier.com/locate/ejim

Original Article

Effectiveness of a Geriatric Care Model for frail older adults in primary care: Results from a stepped wedge cluster randomized trial

Emiel O. Hoogendijk [a,b,c,*], Henriëtte E. van der Horst [a], Peter M. van de Ven [c], Jos W.R. Twisk [c], Dorly J.H. Deeg [c], Dinnus H.M. Frijters [a], Karen M. van Leeuwen [a,d], Jos P.C.M. van Campen [e], Giel Nijpels [a], Aaltje P.D. Jansen [a], Hein P.J. van Hout [a]

[a] Department of General Practice & Elderly Care Medicine, EMGO + Institute for Health and Care Research, VU University Medical Center, Amsterdam, The Netherlands
[b] Gérontopôle, Department of Internal Medicine and Geriatrics, Toulouse University Hospital, Toulouse, France
[c] Department of Epidemiology & Biostatistics, EMGO + Institute for Health and Care Research, VU University Medical Center, Amsterdam, The Netherlands
[d] Department of Health Sciences, EMGO + Institute for Health and Care Research, Faculty of Earth & Life Sciences, VU University, Amsterdam, The Netherlands
[e] Department of Geriatric Medicine, Slotervaart Hospital, Amsterdam, The Netherlands

WORKED EXAMPLE 1: CLOSED COHORT

- My assumptions:
  - 90% power, $\alpha = 0.05$
  - $t = 4$ sequences
  - m=30 individuals per practice (m=24 after applying 20% attrition)
  - Standard deviation = 7.1
  - Target difference = 3
  - wpICC = 0.02, CAC=0.8, IAC=0.66

WORKED EXAMPLE 1: CLOSED COHORT

- Calculate total sample size required under individual randomization: $N_{ind} = 238$
- Calculate design effect due to clustering: $Deff_c = 1 + (m-1)\rho_0 = 1.46$
- Calculate R for cohort design:

$$R = \frac{m\rho_0\omega + (1-\rho_0)\tau}{1+(m-1)\rho_0} = \frac{1.0308}{1.48} = 0.696$$

- Calculate design effect due to time:

$$Deff_t = \frac{3t(1-R)(1+tR)}{(t^2-1)(2+tR)} = \frac{13.8040}{71.76} = 0.1924$$

Don't do it!

- Calculate required number of clusters:

$$k = \frac{N_{ind} \times Deff_c \times Deff_t}{m} = 2.8$$

- Round to a multiple of the number of sequences: 4

WORKED EXAMPLE 1: CLOSED COHORT

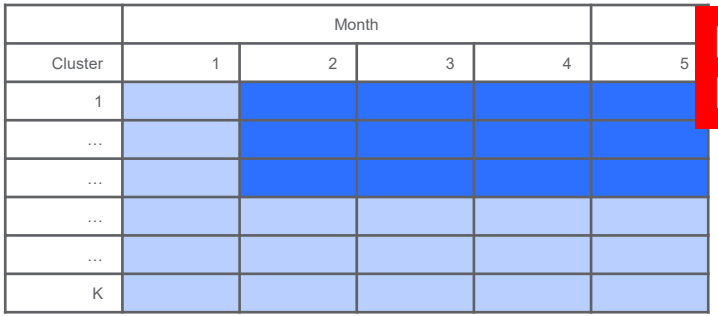
- Let's try again!
- My assumptions:
  - 90% power, $\alpha = 0.05$
  - $t = 4$ sequences
  - m=30 individuals per community each period (m=24 after applying 20% attrition)
  - Standard deviation = 7.1
  - Target difference = 1
  - wpICC = 0.02, CAC=0.8, IAC=0.66

REQUIRED SAMPLE SIZES: SW VS. PARALLEL ARM CRT

Parallel arm

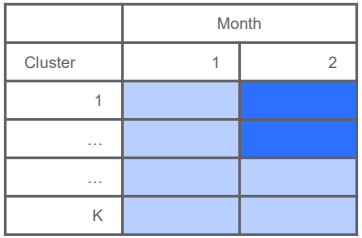
K=130
N=3900

Parallel before & after repeated measures



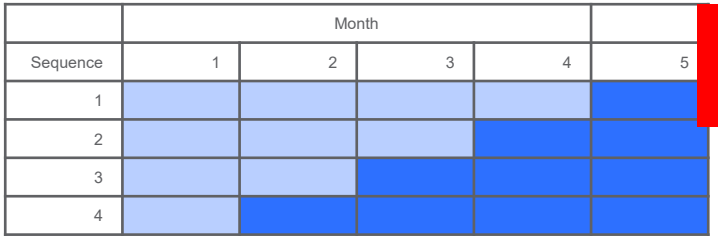
K=38
N=1140

Parallel arm before and after



K=66
N=1980

Stepped wedge



K=28
N=840

NIA IMPACT COLLABORATORY
TRANSFORMING DEMENTIA CARE

JUSTIFICATION FOR THE STEPPED WEDGE DESIGN

- Some reported reasons for using the SW-CRT

1. “A decision has already been made to implement the intervention in a health system”
2. “Clusters reluctant to participate unless offered intervention at some stage during the trial”
3. “I have too few clusters and not enough power for a parallel arm CRT design”
4. “Logistically challenging to implement intervention simultaneously in many clusters”
5. “There is less risk of bias since each cluster serves as their own control”
6. “Ethically inappropriate to withhold a beneficial intervention”
7. “I have always wanted to try a stepped wedge”
8. “It will make my grant more attractive for the funder”

REASON 1: INTERVENTION MUST BE IMPLEMENTED

- YES
  - Decision has been made by stakeholder to implement a program so as to exert its expected benefits
  - SW-CRT design allows more rigorous evaluation than a non-randomized (before and after) design
- NO
  - Will have to convince stakeholders and sponsors of the importance of randomization
  - Will have to reconcile need for adherence to allocated implementation schedule with stakeholder preferences and priorities

REASON 2: TOO DIFFICULT TO RECRUIT

- YES
  - Easier to recruit clusters to the trial if they are offered something “new”
- NO
  - Some clusters may have to wait a very long time and lose interest
  - Intervention may not work or may even be harmful
  - Consider parallel arm design with control clusters offered beneficial intervention at the end of the trial or control clusters offered a reduced version of intervention

REASON 3: TOO FEW CLUSTERS AVAILABLE

- YES
  - The SW design *usually* requires fewer clusters than parallel arm design (ICC or cluster sizes per interval are large)
- NO
  - Check whether power calculations accounted for the CAC
  - A CRT with very few clusters is never a good idea!
  - Consider more efficient parallel arm designs (e.g., before and after CRT)

REASON 4: LOGISTICAL FEASIBILITY

- YES
  - May not have adequate implementation teams for all clusters at the same time

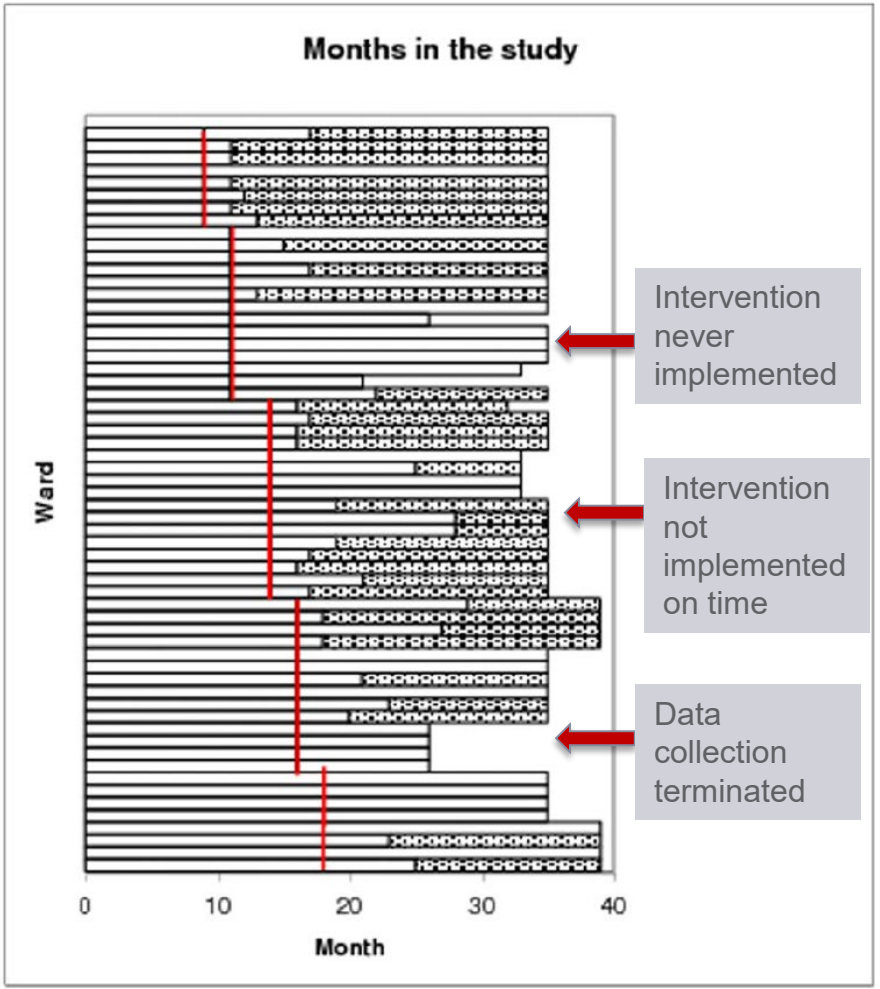
Parallel CRT with staggered implementation

	Time			
Hospitals	6	12	18	24
1				
2				
3				
4				
5				
6				

- NO
  - SW design can bring new logistical challenges, e.g., need to have all IRB approvals in place at the start, challenges in adhering to implementation schedule
  - Alternative: consider parallel arm design with staggered implementation

EXAMPLE

- The Feedback Intervention Trial – Improving Hand Hygiene Compliance in UK Healthcare Workers (Fuller ea, 2012)



REASON 5: TO REDUCE BIAS

- YES
  - It is partially true that each cluster serves as their own control
- NO
  - Intervention is confounded with time by design and appropriate modeling of the time effect can be difficult
  - SW-CRT can introduce additional risks of bias (e.g., contamination, time-varying effects, attrition)

REASON 6: ETHICAL REQUIREMENT

- YES
  - None
- NO
  - Requirement for equipoise still applies
  - No ethical justification for delaying intervention to some clusters
  - All clusters, but not necessarily all *participants* will receive intervention

SUMMARY

- SW-CRT is a novel design enthusiastically embraced by trialists
- Methodology is still evolving
- Intervention confounded with time *by design* – necessarily need a model-based analysis
- Subject to several risks of bias
- While it can be a good choice in some circumstances, we ought to think carefully before adopting it

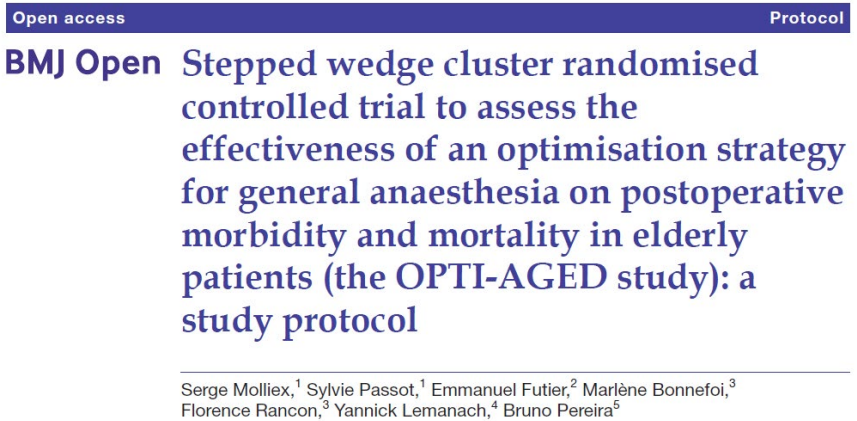
KEY REFERENCES

- Hussey MA & Hughes JP (2007) Design and analysis of stepped wedge cluster randomized trials. *Contemporary Clinical Trials;* 28:182-191
- Hughes JP, Granston TS, Heagerty PJ. (2015) On the design and analysis of stepped wedge trials. *Contemporary Clinical Trials*. 45(Pt A):55-60
- Hooper R, Bourke L (2015) Cluster randomised trials with repeated cross sections: alternatives to parallel group designs. *BMJ*. Jun 8;350:h2925
- Hemming K, Lilford R, Girling AJ. (2015) Stepped-wedge cluster randomised controlled trials: a generic framework including parallel and multiple level designs. *Statist. Med;* 34(2):181-196
- Copas AJ e.a. (2015) Designing a stepped wedge trial: three main designs, carry-over effects and randomisation approaches. *Trials*; 16:352
- Girling AJ and Hemming K. (2016) Statistical efficiency and optimal design for stepped cluster studies under linear mixed effects models. *Statist Med* 35(13):2149-66
- Hooper R, Teerenstra S, de Hoop E, Eldridge S. (2016) Sample size calculation for stepped wedge and other longitudinal cluster randomised trials. *Statist Med* 35(26):4718-4728
- Kasza J, Hemming K, Hooper R, Matthews JNS, Forbes AB et al. (2017) Impact of non-uniform correlation structure on sample size and power in multiple-period cluster randomised trials. *Stat Methods Med Res.*
- Hemming K, Taljaard M, McKenzie JE, et al. Reporting of stepped wedge cluster randomised trials: Extension of the CONSORT 2010 statement with explanation and elaboration. *BMJ*. 2018;363:k1614
- Kasza J, Forbes AB. Information Content of Cluster–Period Cells in Stepped Wedge Trials. *Biometrics* 2019; 75, 144-152

SUPPLEMENTARY SLIDES

WORKED EXAMPLE 2: CONTINUOUS RECRUITMENT

- Sample size parameters (as stated in manuscript):
  - 90% power, $\alpha = 0.05$
  - $t = 5$ sequences
  - m=? individuals per hospital per period
  - Control proportion = 0.24
  - Target difference = 0.072
  - “ICC = 0.005-0.05”
  - CAC=?



WORKED EXAMPLE 2: CONTINUOUS RECRUITMENT

- My assumptions:
  - 90% power, $\alpha = 0.05$
  - $t = 5$ sequences
  - m=90 patients per hospital per period
  - Control proportion = 0.24
  - Target difference = 0.072
  - wpICC = 0.01
  - CAC=0.8

WORKED EXAMPLE 2: CONTINUOUS RECRUITMENT

- Calculate total sample size required under individual randomization: $N_{ind} = 1314$
- Calculate design effect due to clustering: $Deff_c = 1 + (m-1)\rho_0 = 5.45$
- Calculate R for cross-sectional design:

$$R = \frac{m\rho_0\omega}{1+(m-1)\rho_0} = \frac{3.6}{5.45} = 0.661$$

- Calculate design effect due to time:

$$Deff_t = \frac{3t(1-R)(1+tR)}{(t^2-1)(2+tR)} = \frac{21.8909}{127.32} = 0.1719$$

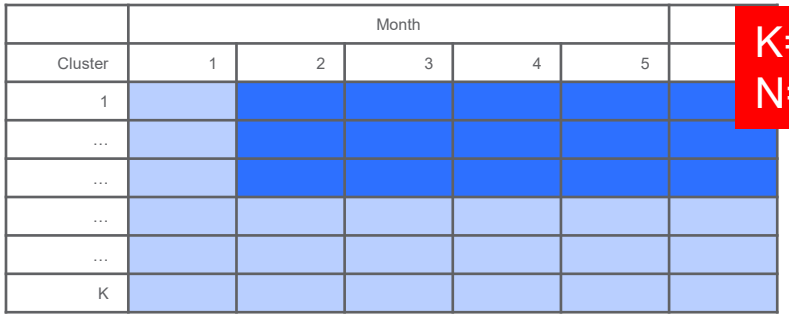
- Calculate required number of clusters: $k = \frac{N_{ind} \times Deff_c \times Deff_t}{m} = 13.7$
- Round to a multiple of the number of sequences: 15

EXAMPLE 2: COMPARE SAMPLE SIZES, SW VS. PARALLEL

Parallel arm

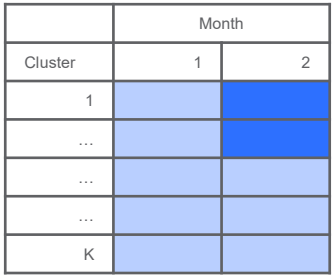
K=94
N=8460

Parallel before & after repeated measures



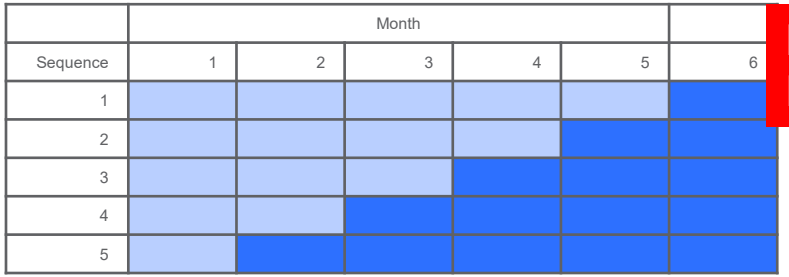
K=24
N=12,960

Parallel arm before and after



K=46
N=8280

Stepped wedge



K=15
N=8100

NIA IMPACT COLLABORATORY
TRANSFORMING DEMENTIA CARE

SAMPLE SIZE RESOURCES

- R package ‘swCRTdesign’ http://faculty.washington.edu/jphughes/pubs.html
  - Allows for fractional treatment indicator, incomplete designs, cluster treatment heterogeneity (but not correlation decay)
- R-Shiny (Hemming & Kasza) https://clusterrcts.shinyapps.io/rshinyapp/
  - Includes parallel arm longitudinal, stepped wedge, and cross-over designs
  - Continuous, binary or count outcomes
  - Repeated cross-sectional and cohort designs
  - Equal or unequal allocation
  - Complete or incomplete designs (but not fractional treatment indicator)
  - Adjustments for cluster size variability
  - Allows for correlation decay and cluster treatment heterogeneity


NIA IMPACT
COLLABORATORY
TRANSFORMING DEMENTIA CARE