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

DESIGNING TRIALS WITHIN IMPLEMENTATION LABORATORIES: STANDARD AND LONGITUDINAL CLUSTER RANDOMIZED DESIGNS

SOCIETY FOR CLINICAL TRIALS ANNUAL MEETING

Portland May, 2018

MONICA TALJAARD

SENIOR SCIENTIST, CLINICAL EPIDEMIOLOGY PROGRAM, OHRI
ASSOCIATE PROFESSOR, SCHOOL OF EPIDEMIOLOGY AND PUBLIC HEALTH, UNIVERSITY OF OTTAWA

OUTLINE

1. Key characteristics of implementation laboratories and design implications
2. Example: The RAPiD trial
3. Sample size calculation methods for longitudinal cluster randomized designs
4. Application to the RAPiD trial
5. Conclusions

KEY CHARACTERISTICS OF IMPLEMENTATION LABORATORIES

1. Complex interventions → necessarily require cluster randomization
2. Embedded within existing, large-scale initiatives → potentially large number of clusters available
3. Interventions targeted at the health provider or health system → no need for patient consent
4. Outcomes routinely collected → large cluster sizes available; availability of retrospective (pre-intervention) data
5. Not “intervention vs no intervention” but “how to do it better” → multiple intervention arms with small effects

EXAMPLE: RAPiD

RAPID: REDUCING ANTIBIOTIC PRESCRIBING IN DENTISTRY

PLOS | MEDICINE

RESEARCH ARTICLE

An Audit and Feedback Intervention for Reducing Antibiotic Prescribing in General Dental Practice: The RAPiD Cluster Randomised Controlled Trial

Paula Elouafkaoui[1,2], Linda Young[1]*, Rumana Newlands[3], Eilidh M. Duncan[3], Andrew Elders[4], Jan E. Clarkson[1,2], Craig R. Ramsay[3], Translation Research in a Dental Setting (TRiaDS) Research Methodology Group[¶]

1 NHS Education for Scotland (NES), Dundee Dental Education Centre, Frankland Building, Dundee, United Kingdom, **2** Dental Health Services Research Unit (DHSRU), University of Dundee, Park Place, Dundee, United Kingdom, **3** Health Services Research Unit (HSRU), University of Aberdeen, Health Sciences Building, Foresterhill, Aberdeen, United Kingdom, **4** NMAHP Research Unit, Glasgow Caledonian University, Cowcaddens Road, Glasgow, United Kingdom

¶ Membership of the Translation Research in a Dental Setting (TRiaDS) Research Methodology Group is provided in the Acknowledgments.
* linda.young@nes.scot.nhs.uk

OPEN ACCESS

Citation: Elouafkaoui P, Young L, Newlands R, Duncan EM, Elders A, Clarkson JE, et al. (2016) An Audit and Feedback Intervention for Reducing Antibiotic Prescribing in General Dental Practice: The RAPiD Cluster Randomised Controlled Trial. PLoS Med 13(8): e1002115. doi:10.1371/journal.pmed.1002115

Abstract

Background

EXAMPLE: RAPiD

- **Objective:** To reduce inappropriate prescribing of antibiotics by dentists
- **Interventions**: Individualised graphical Audit and Feedback (A&F) versus individualised graphical A&F plus written behaviour change message
- **Design:** Cluster randomized trial
- **Primary outcome**: Total number of antibiotic items dispensed per 100 NHS treatment claims over the 12 month post-intervention period
- **Participants**: All 795 NHS general dental practices in Scotland
- **Results**: A&F plus written messages more effective than A&F alone

EXAMPLE: RAPiD

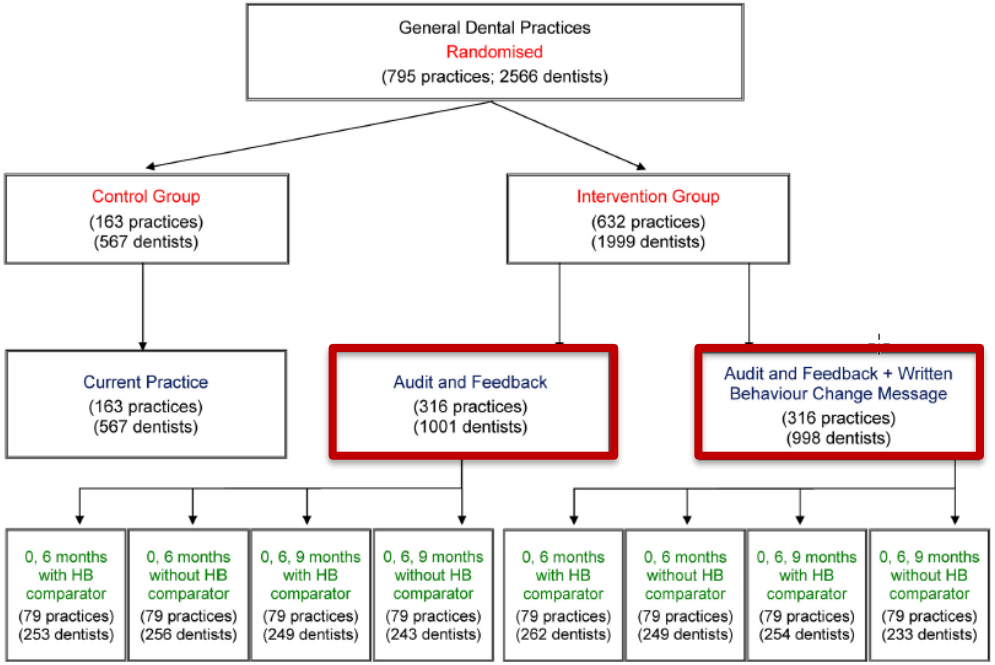
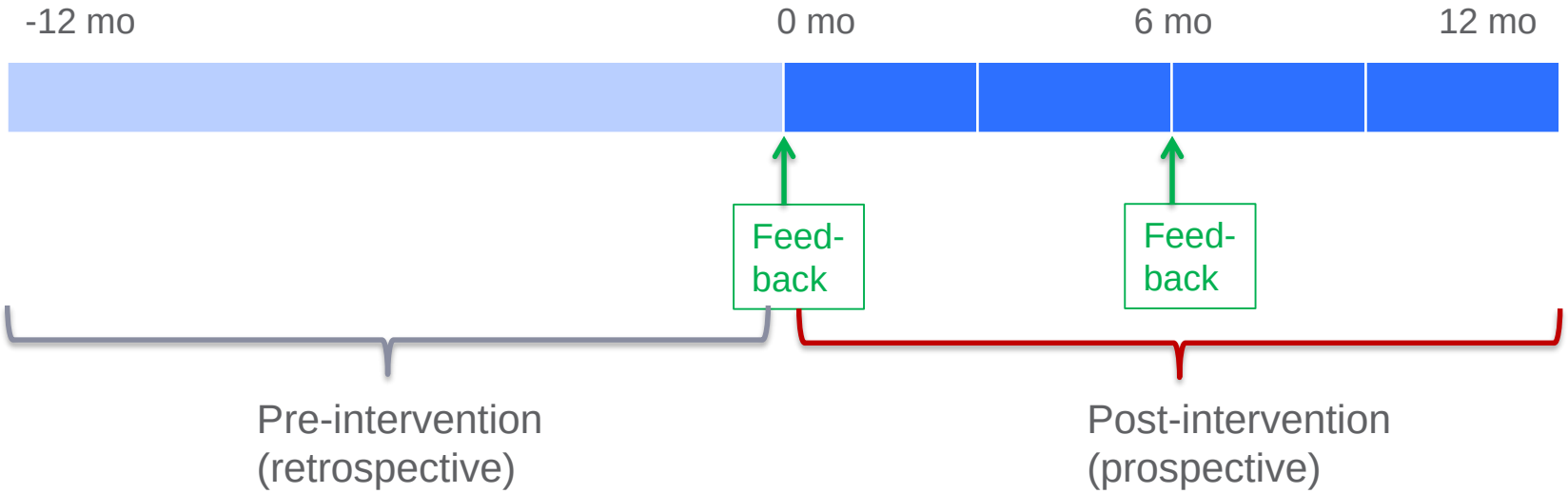


Fig 1. RAPiD trial design. Description of interventions: Audit and feedback—a line graph plotting an individual dentist's monthly antibiotic prescribing rate. Written behaviour change message—text added below the individualised line graph synthesising and reiterating national guidance recommendations for antibiotic prescribing. HB comparator—addition of a line to the individualised line graph plotting the monthly antibiotic prescribing rate of all dentists in that dentist's health board. 0, 6 months—allocated intervention delivered at months 0 and 6. 0, 6, 9 months—allocated intervention delivered at months 0, 6, and 9.

INTERVENTION AND DATA COLLECTION SCHEDULE



SAMPLE SIZE CALCULATION (CLUSTER- LEVEL)

- Effect size: Mean absolute difference of 0.85% in prescription rates measured at practice level (10% relative difference)
- Standard deviation = 8%
- Correlation with mean prescription rate at baseline = 0.9
- 80% power
- Two-sided significance level 2.5%

- Using ANCOVA of cluster means, required number of practices = 316/arm

ANALOGOUS CALCULATION (INDIVIDUAL-LEVEL)

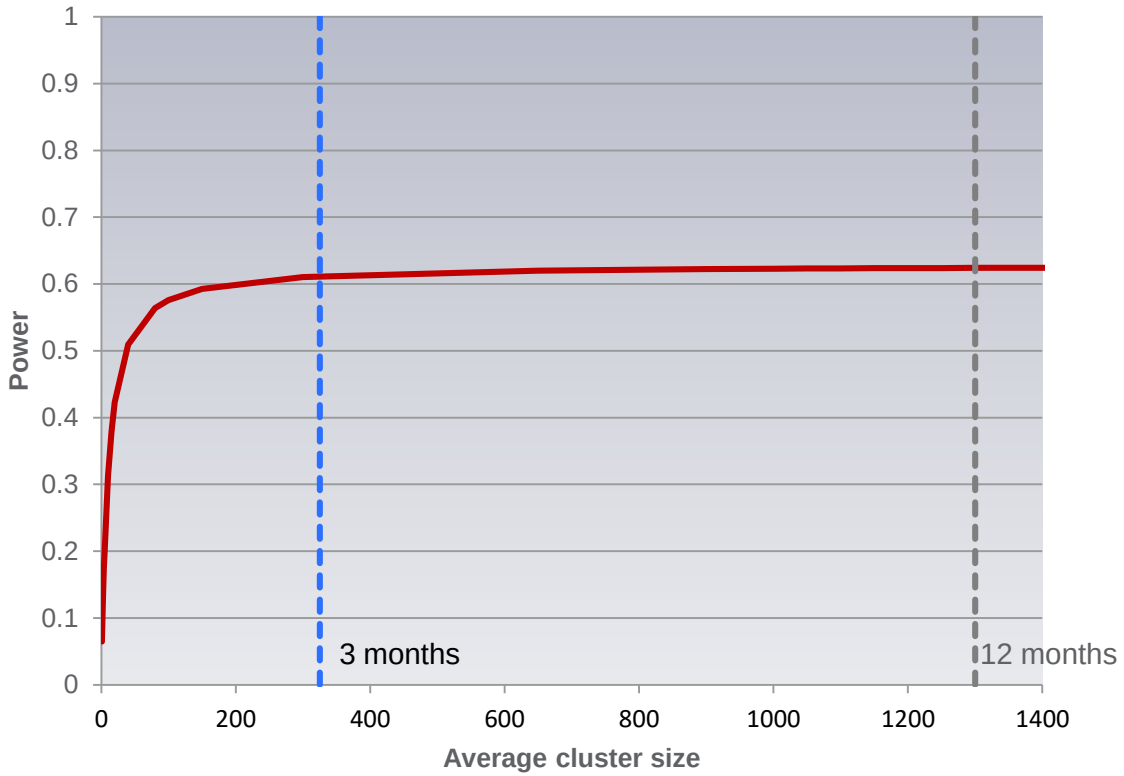
- Effect size: Absolute difference of 0.0085 in proportions (10% relative)
- Control arm proportion = 0.085
- Correlation with mean prescription rate at baseline =0.9
- Average cluster size: 1300 (over 12 months)
- Intracluster Correlation Coefficient (ICC) ~0.08

- Using individual-level ANCOVA, required number of practices = 322/ arm

POWER CURVE FOR RAPID

- Power to detect a 10% relative difference with 322 practices per arm

- My question: Is there an alternative design that would have allowed making better use of the data?



METHODS FOR LONGITUDINAL CRT DESIGNS

LONGITUDINAL CRT DESIGNS

- Substantial recent attention to methodology for longitudinal cluster randomized designs
  - Repeated measurements of the outcome over time
- Clusters could be exposed to the same condition over time (e.g. parallel arm design), or cross between conditions (e.g. parallel before and after, stepped wedge)
- Outcomes could be observed on the same individuals over time (cohort design) or different individuals (cross-sectional design)

EXAMPLE PARALLEL ARM DESIGNS

A. Simple parallel arm

	Time
Cluster	12 months
1	
…	
…	
…	
…	
K	

Intervention

Control

B. Parallel arm repeated measures

	Time					
Cluster	2	4	5	6	10	12
1						
…						
…						
…						
…						
K						

C. Parallel arm before and after (ACTUAL DESIGN)

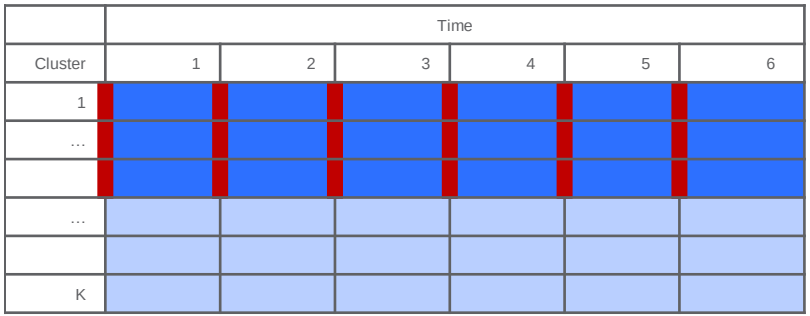
	Time	
Cluster	-12	12
1		
…		
…		
…		
…		
K		

D. Parallel arm before & after repeated measures

	Time											
Cluster	-12	-10	-8	-6	-4	-2	2	4	6	8	10	12
1												
…												
…												
…												
…												
K												

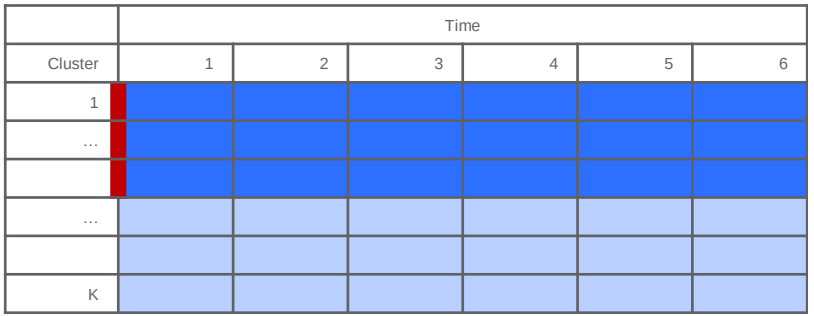
HOW IS INTERVENTION EXPECTED TO WORK?

Repeated delivery



Intervention has an immediate effect but needs repetition to sustain its effects

Single delivery

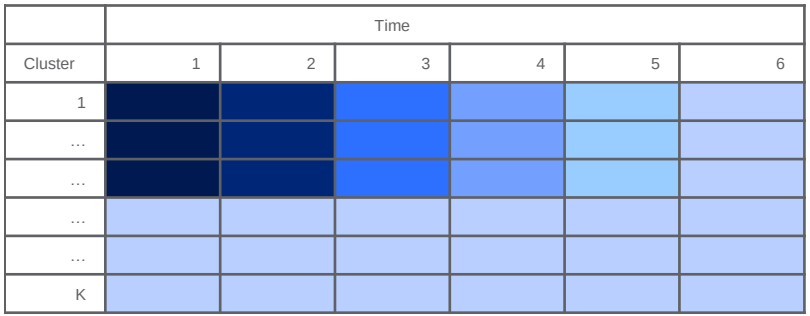


Intervention has an immediate effect which is sustained

Timing for assessing intervention effect: Average across post-intervention period

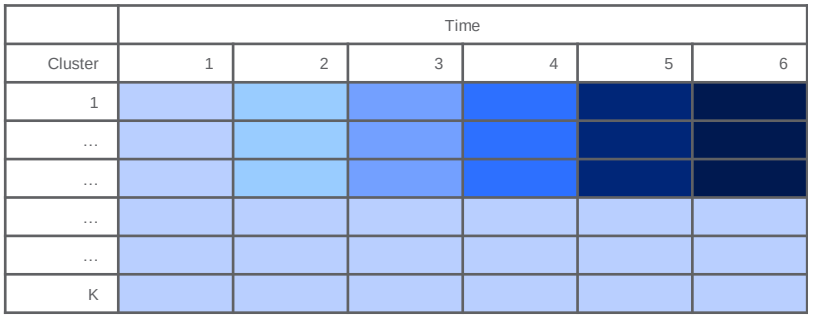
HOW IS INTERVENTION EXPECTED TO WORK?

Immediate effect with a subsequent decay



Intervention has an immediate effect which decays over time

Gradual effect



Intervention has a gradual effect

Timing for assessing intervention effect: Average effect across post-intervention period not meaningful

METHODS FOR LONGITUDINAL CRTs

- Hooper et al (2015, 2016) present methodology for longitudinal CRTs when time-averaged effect is of interest
  - Underlying analytical model
  - Sample size calculation matching the analytical model
- Their approach allows for
  - Cross-sectional and cohort designs
  - Any type of longitudinal design (parallel, stepped wedge, cross-over)

- 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

UNDERLYING ANALYTICAL MODEL

- Linear mixed effects model for repeated cross-sectional design:

fixed

$$Y_{ijl} = \mu + \beta_j + \theta X_{ij} + u_i + (u\tau)_{ij} + e_{ijl}; \quad \text{random}$$

$$u_i \sim N\left(0, \sigma_u^2\right); \ (u\tau)_{ij} \sim N\left(0, \sigma_{u\tau}^2\right); \qquad e_{ijl} \sim N(0, \sigma_e^2)$$

β_j = fixed effect of time

X_{ij} = 1 if intervention, 0 if control

θ = intervention effect

u_i = random effect for cluster

$(u\tau)_{ij}$ = random time effect for cluster

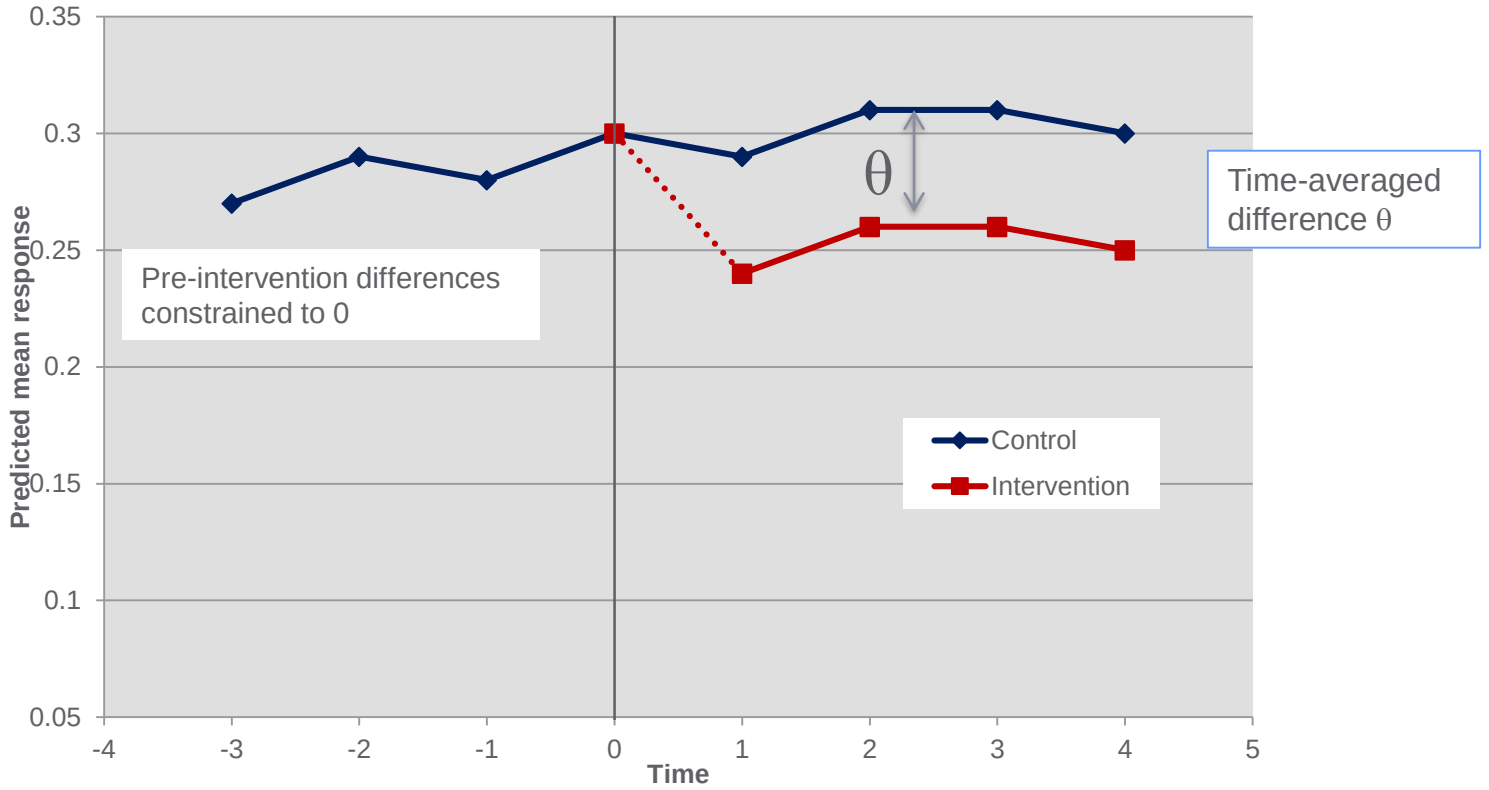
e_{ijl} = residual

$i = 1, \text{K}, k$ clusters

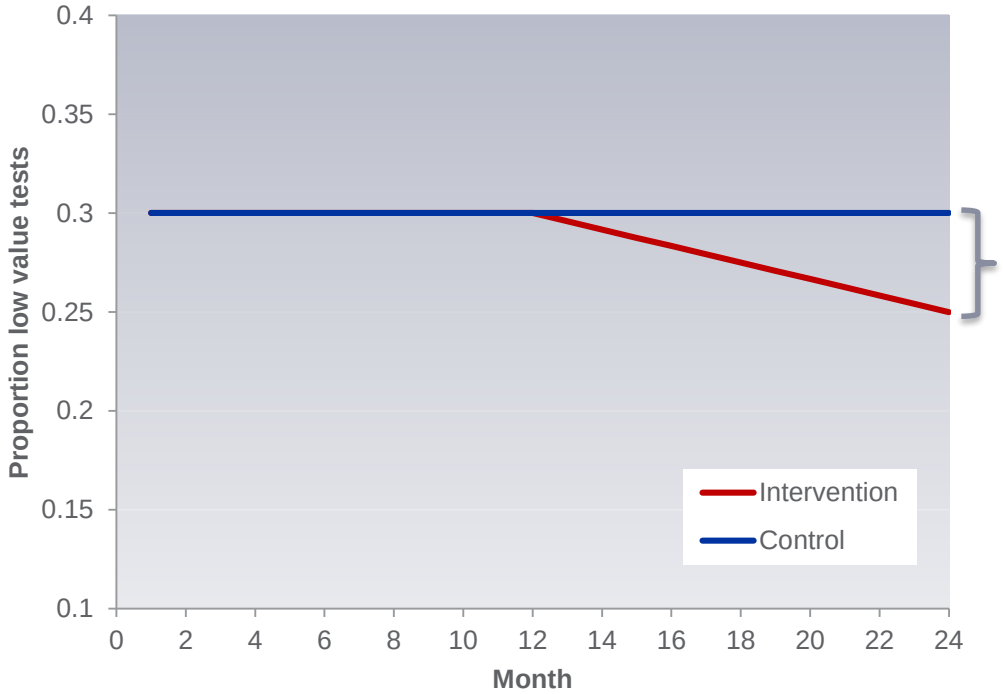
$j = 1, \text{K}, T$ periods

$l = 1, \text{K}, m$ individuals

ASSUMED INTERVENTION EFFECT



(DIFFERENT THAN GRADUAL CHANGE)



Difference at trial end, or slope change

RANDOM EFFECTS ASSUMPTIONS

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

$$wpICC = \rho_0 = \frac{\sigma_u^2 + \sigma_{u\tau}^2}{\sigma_u^2 + \sigma_{u\tau}^2 + \sigma_e^2};$$

- Between-period ICC: between two individuals in the same cluster but different periods

$$bpICC = \rho_1 = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_{u\tau}^2 + \sigma_e^2}$$

- Note: bpICC ≤ wpICC

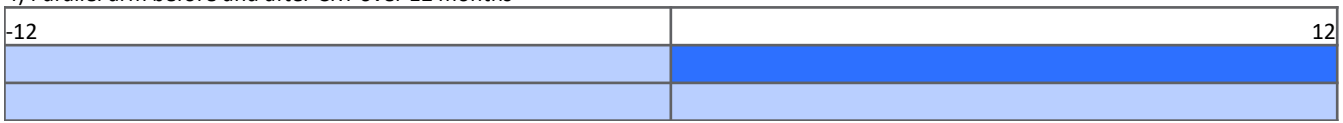
EXAMPLE: THE RAPID TRIAL

- Illustrate implications of various longitudinal designs for the required number of practices to detect a 10% relative difference in RAPiD
- Assume a within-period ICC of 0.08 (over 1 year period)
- Assume bpICC = 0.9*wpICC

REQUIRED SAMPLE SIZES PARALLEL ARM DESIGNS

Actual design

4) Parallel arm before and after CRT over 12 months



644 practices
1,674,400 visits

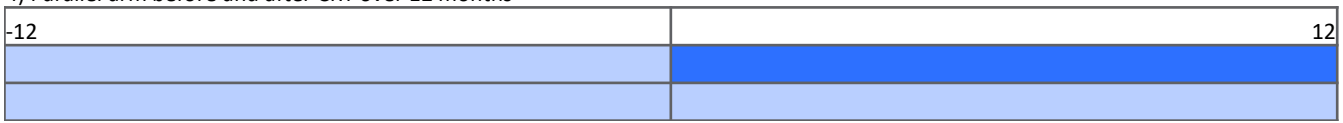
REQUIRED SAMPLE SIZES PARALLEL ARM DESIGNS

1) Parallel arm CRT over 3 months

3472 practices
1,128,400 visits

Actual design

4) Parallel arm before and after CRT over 12 months



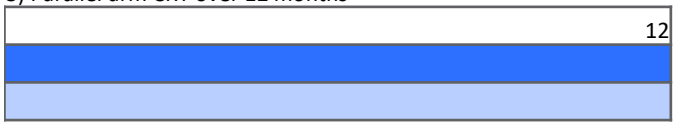
644 practices
1,674,400 visits

REQUIRED SAMPLE SIZES PARALLEL ARM DESIGNS

1) Parallel arm CRT over 3 months

3472 practices
1,128,400 visits

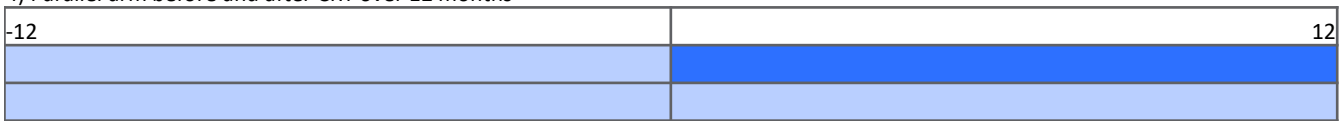
3) Parallel arm CRT over 12 months



3154 practices
4,100,200 visits

Actual design

4) Parallel arm before and after CRT over 12 months



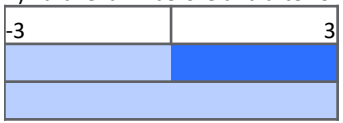
644 practices
1,674,400 visits

REQUIRED SAMPLE SIZES PARALLEL ARM DESIGNS

1) Parallel arm CRT over 3 months

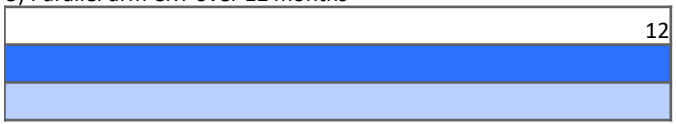
3472 practices
1,128,400 visits

2) Parallel arm before and after CRT over 3 months



836 practices
543,400 visits

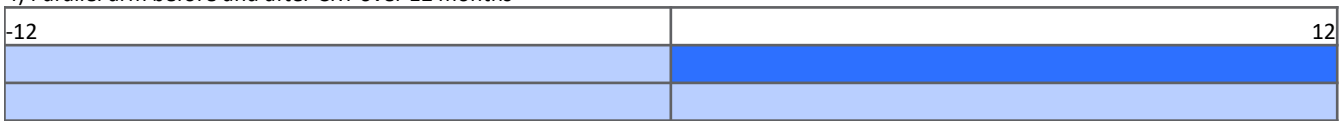
3) Parallel arm CRT over 12 months



3154 practices
4,100,200 visits

Actual design

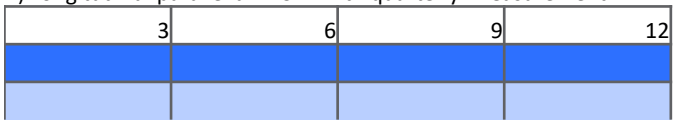
4) Parallel arm before and after CRT over 12 months



644 practices
1,674,400 visits

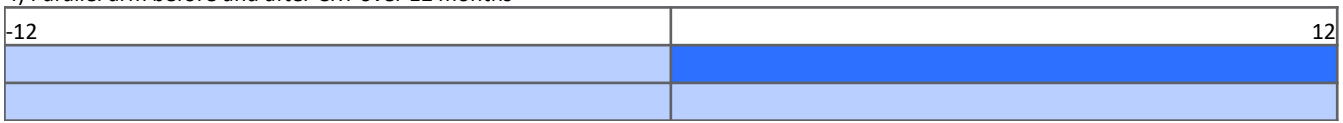
REQUIRED SAMPLE SIZES PARALLEL ARM DESIGNS

1) Longitudinal parallel arm CRT with quarterly measurement



3138 practices
4,079,400 visits

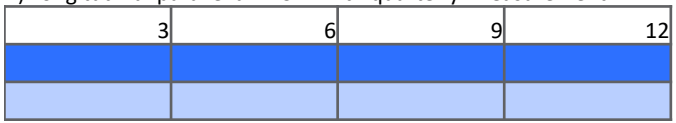
4) Parallel arm before and after CRT over 12 months



644 practices
1,674,400 visits

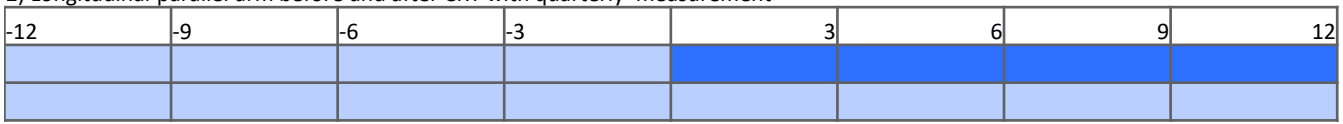
REQUIRED SAMPLE SIZES PARALLEL ARM DESIGNS

1) Longitudinal parallel arm CRT with quarterly measurement



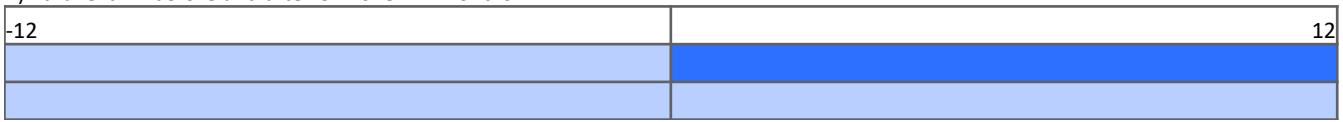
3138 practices
4,079,400 visits

2) Longitudinal parallel arm before and after CRT with quarterly measurement



220 practices
572,000 visits

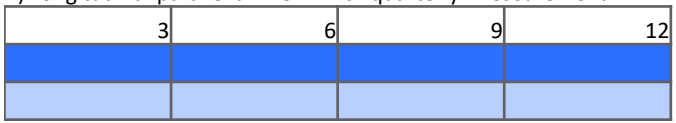
4) Parallel arm before and after CRT over 12 months



644 practices
1,674,400 visits

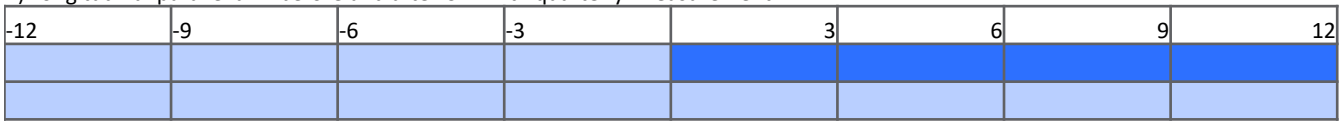
REQUIRED SAMPLE SIZES PARALLEL ARM DESIGNS

1) Longitudinal parallel arm CRT with quarterly measurement



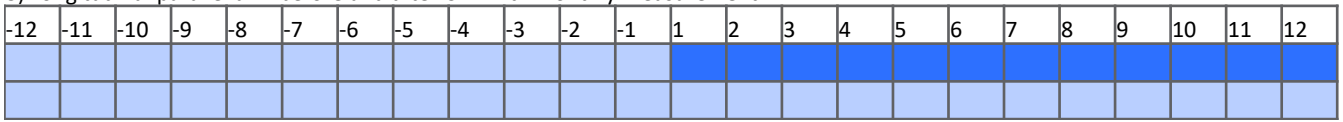
3138 practices
4,079,400 visits

2) Longitudinal parallel arm before and after CRT with quarterly measurement



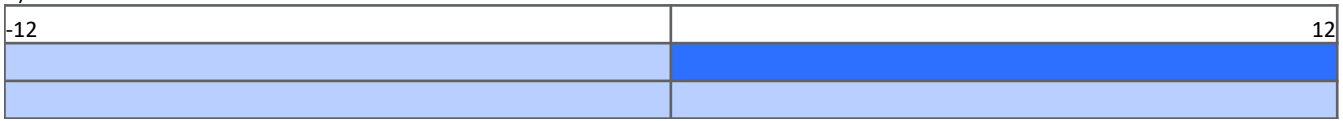
220 practices
572,000 visits

3) Longitudinal parallel arm before and after CRT with monthly measurement



112 practices
290,304 visits

4) Parallel arm before and after CRT over 12 months



644 practices
1,674,400 visits

CONCLUSIONS

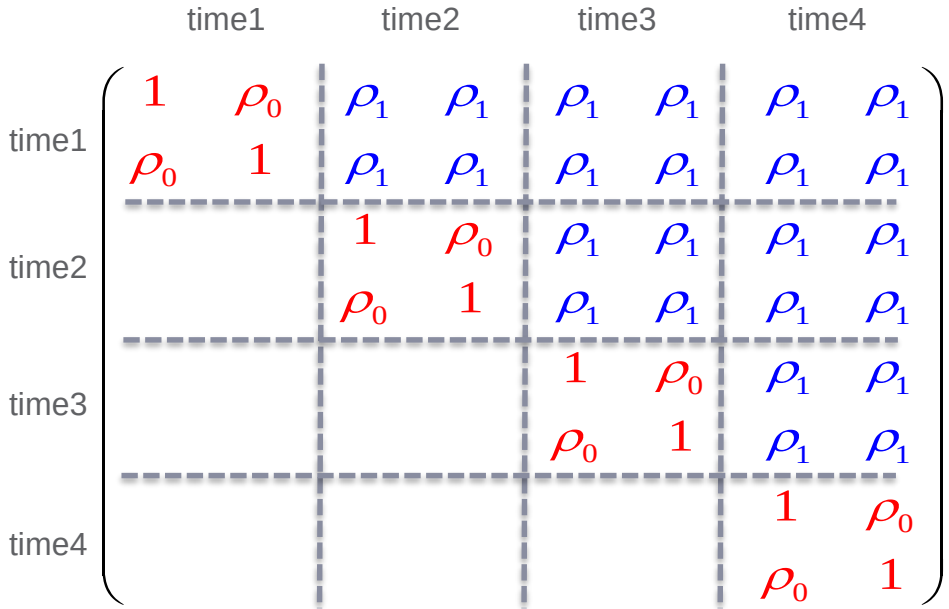
- Recommendations:
  - Inspect power curve at the design stage
  - If design is inefficient, consider adopting pre and post repeated measures
  - May decrease required number of clusters (or facilitate ability to power for important subgroup differences and smaller effects in multi-arm trials)

CONCLUSIONS

- Concerns:
  - Is a time-averaged difference appropriate (may need to account for learning and decay effects)
  - Need reliable data to inform sample size calculation (within-period and between-period ICCs)
  - Unduly restricting the number of clusters is risky
  - Modeling time as a categorical variable may not be ideal with many time intervals

RANDOM EFFECTS ASSUMPTIONS

- Within-period and between-period ICCs



wpICC

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

bpICC

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

"CLUSTER AUTOCORRELATION COEFFICIENT"

- The ratio of the between-period and within-period ICCs is called the "Cluster Autocorrelation Coefficient" (CAC)

$$CAC = \frac{bpICC}{wpICC} = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_{u\tau}^2} = \pi$$

- Rather than specifying a within-period and a between-period ICC, it is easier to specify the within-period ICC and then specify the CAC
- CAC measures the reduction in bpICC relative to wpICC
- In the absence of a prior estimate, a general rule of thumb is CAC between 0.6 to 0.8

SAMPLE SIZE CALCULATION PROCEDURE

- Involves five steps:

1. Calculate **total** required sample size under individual randomization $N=n_1+n_2$
2. Multiply by design effect due to clustering $Deff_c = 1+(m-1)\rho_0$
3. Multiply by design effect due to repeated measures $Deff_t$ (depends on type of longitudinal design - see next slide)
4. Divide by cluster size per period (m) to determine **total** required number of clusters (K)

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

DESIGN EFFECT DUE TO REPEATED ASSESSMENT

- *Deff*$_t$ depends on the type of longitudinal design, the number of pre and post repeated measurements and r = the “cluster mean correlation”

Design	Deff$_t$
Parallel (1 pre and 1 post)	$(1-r^2)$
Parallel (*u* pre and *v* post)	$\frac{(1-r)\left[1+(u+v-1)r\right]}{v\left[1+(u-1)r\right]}$

CLUSTER MEAN CORRELATIONS

- The cluster mean correlation (*r*) depends on cluster size (*m*), ICC and CAC
  - Cluster mean correlation for cross-sectional designs:

$$r = \frac{m\rho_0\pi}{1+(m-1)\rho_0}$$

  - Cluster mean correlation for cohort designs:

$$r = \frac{m\rho_0\pi + (1-\rho_0)\tau}{1+(m-1)\rho_0}$$

where ρ_0 is the within-period ICC, π is the Cluster Autocorrelation Coefficient (CAC) and τ is the Individual Autocorrelation Coefficient (IAC)