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

INNOVATIVE DESIGNS TO EVALUATE IMPLEMENTATION STRATEGIES

CENTER FOR IMPLEMENTATION RESEARCH LAUNCH

January 19, 2018

MONICA TALJAARD

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

The Ottawa Hospital | L'Hôpital d'Ottawa

OUTLINE

1. Review three conventional study designs that can be used to evaluate implementation strategies
2. Introduce a novel design: the *Stepped Wedge*
3. Example: The pre-op testing trial
4. Conclusions

CONVENTIONAL STUDY DESIGNS

- Concerned here with study designs which can be used to evaluate the effectiveness of *implementation strategies*
- Tools and strategies to help providers / healthcare systems implement evidence-based practices
- Such strategies often involve
  - multiple interacting components;
  - targeted at the health system or provider;
  - outcomes may be assessed on patients.

EXAMPLE: THE PRE-OP TESTING TRIAL

- **Background:** A 2017 report by *Canadian Institute for Health Information* and *Choosing Wisely Canada* found that 36% of low risk surgical patients receive unnecessary medical tests before their surgery (e.g., chest X-rays). The newly established *Choosing Wisely Canada Implementation Research Network* will develop an implementation strategy to help change clinicians' test ordering practices.
- **Intervention***:* Includes an engagement and education program targeting anesthesiologists, surgeons, nurses and administrative staff; as well as tools to overcome possible barriers to practice change.
- **Objective:** To reduce the proportion of low risk patients receiving any unnecessary tests within 60 days before surgery, as recorded in the hospital databases

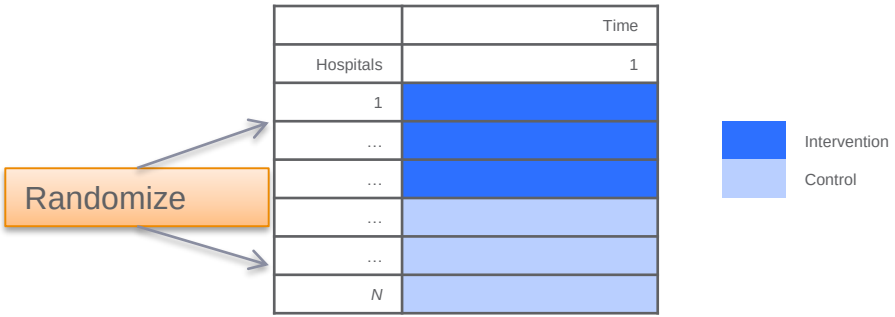
CHOOSING A STUDY DESIGN

- General principle:
  - Choose the most robust design possible to minimize bias while maximizing generalizability
- Minimizing bias (internal validity)
  - Is the observed change in outcome actually caused by the intervention?
- Maximizing generalizability (external validity)
  - Does the finding apply to other hospitals and patients?

RANDOMIZED CONTROLLED TRIALS

- “Gold standard” study design
- Allocate an adequate number of units to either intervention or control arms using a random procedure
- Differences observed between the arms at the end of the study are attributed to the intervention
- Random allocation allows us to rule out alternative explanations for the difference
- Complex interventions (targeted at the entire hospital) necessarily require allocation of *hospitals*, rather than individual *patients*
- This type of study is called a *Cluster randomized trial* (CRT)

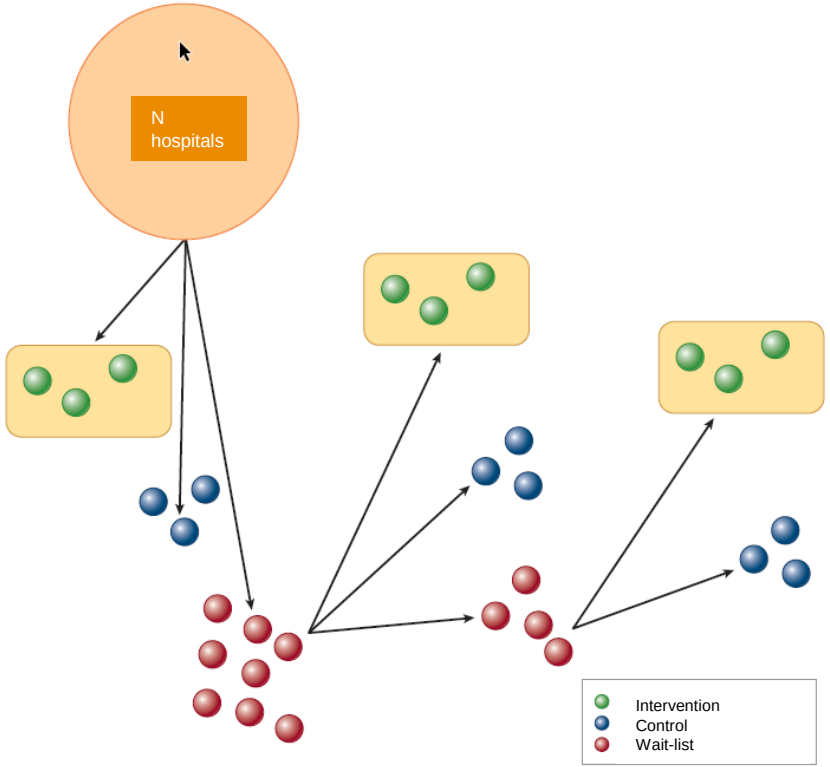
DESIGN OPTION 1: PARALLEL ARM CRT



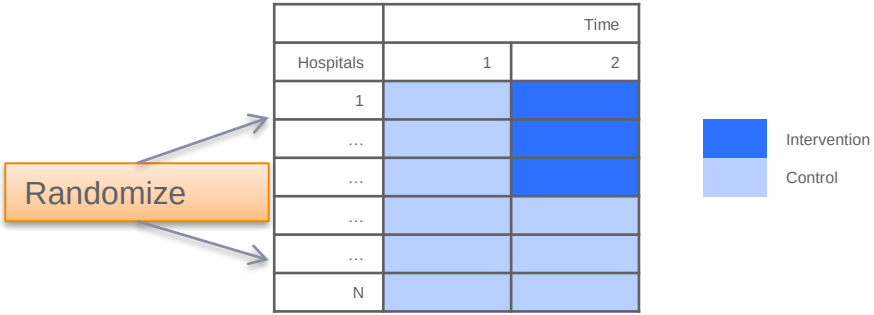
- Recruit *N* (number of) hospitals to participate in the study
- Allocate them to intervention or control arms using a random procedure
- Hospitals in intervention arm implement the intervention; those in control arm continue as usual
- After the intervention, outcomes are observed in both arms

DESIGN OPTION 1: PARALLEL ARM CRT

- Simultaneous implementation of the intervention at many hospitals may be logistically challenging
- An alternative is to randomly allocate hospitals “in waves”



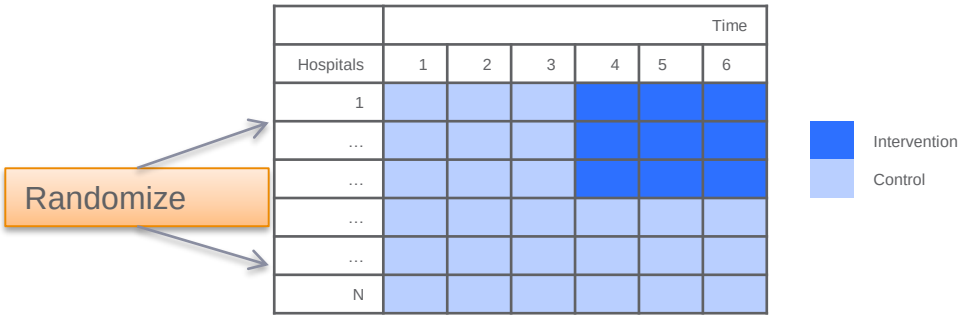
DESIGN OPTION 2: BEFORE AND AFTER PARALLEL ARM CRT



- Add a pre-intervention measurement in both arms

- Advantages:
  - Increases “power” (ability to detect a difference)
  - Can assess whether hospitals in intervention and control arms are similar before intervention
  - Can assess whether hospitals who drop out are similar to those who do not

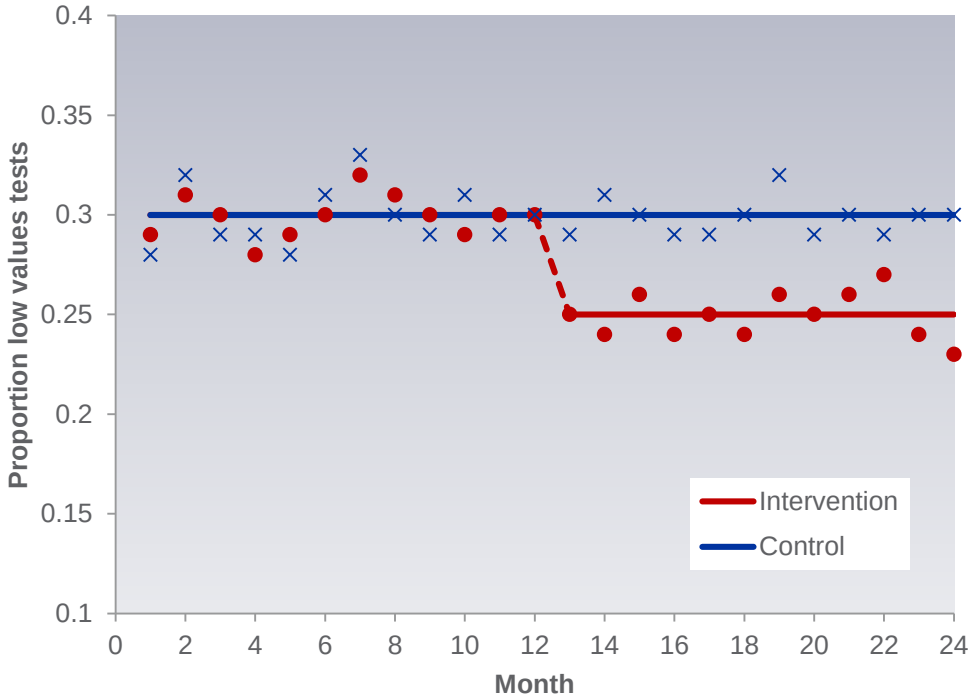
DESIGN OPTION 3: LONGITUDINAL PARALLEL ARM CRT



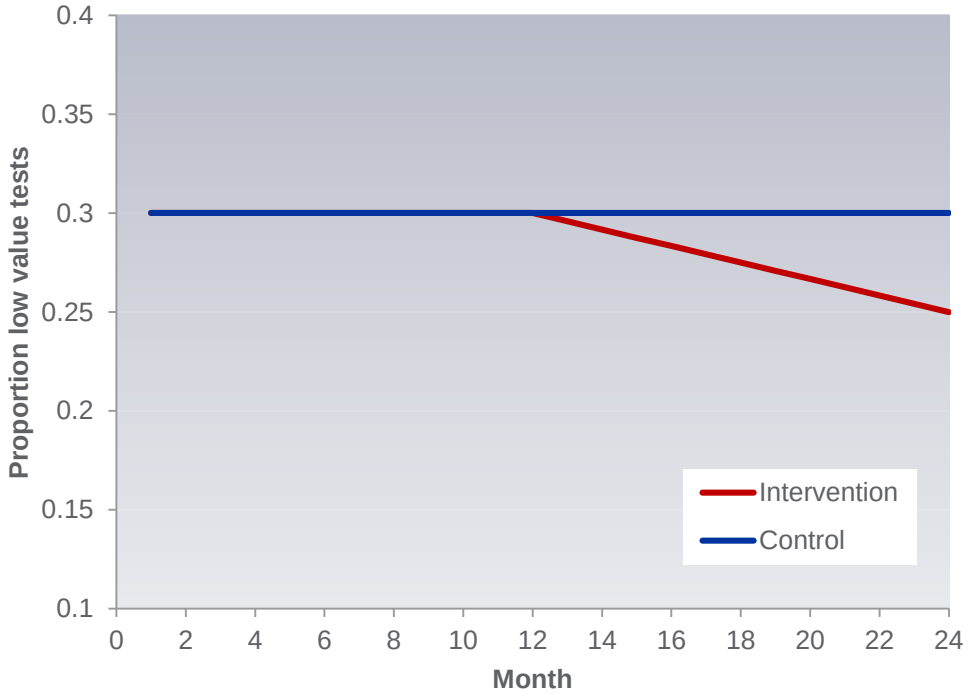
- Multiple observations taken before and after intervention

- Advantages:
  - Increase “power” (ability to detect a difference)
  - Can study how outcomes change over time in response to intervention (immediate, gradual)
  - Can assess whether changes are sustained in the long-run
  - Can assess for presence of “secular trends” (improvements happening naturally over time)

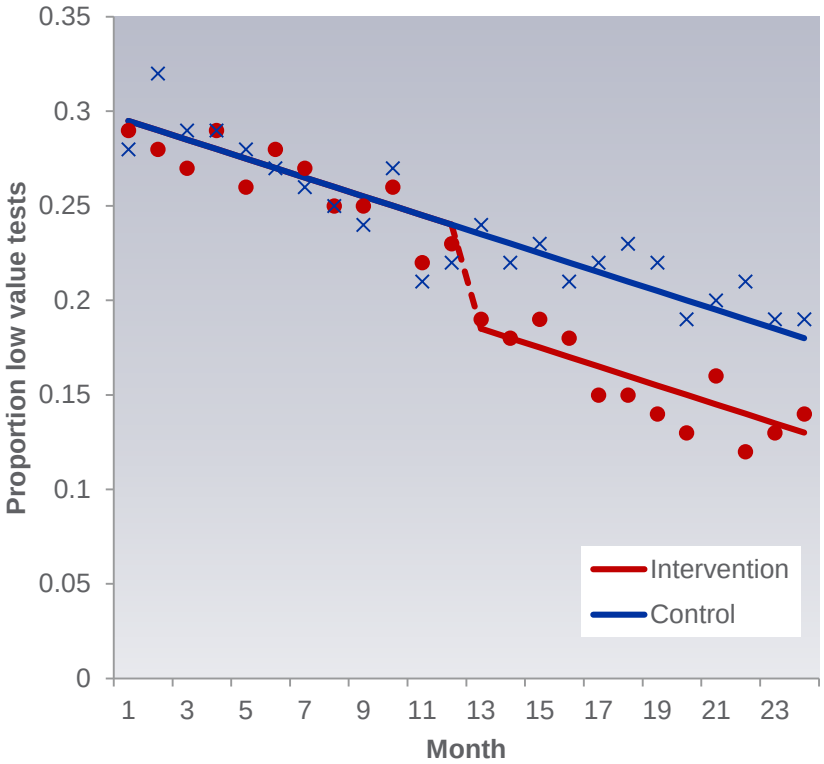
IMMEDIATE CHANGE THAT PERSISTS THROUGH TIME



GRADUAL CHANGE

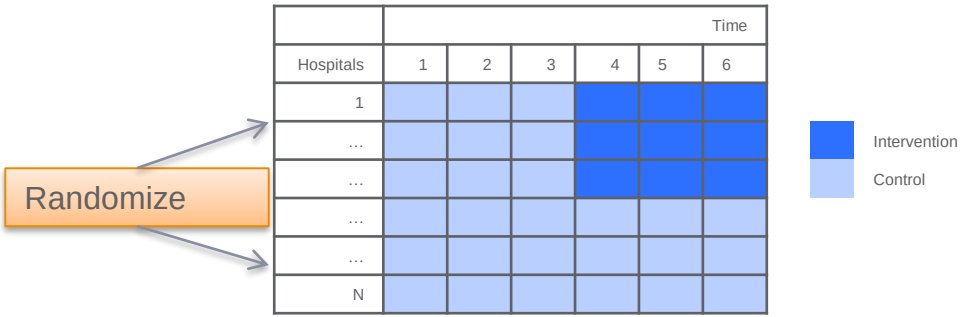


SECULAR TREND



- Test ordering practices already improving even before intervention
- Intervention has an additional effect over and above the secular trend

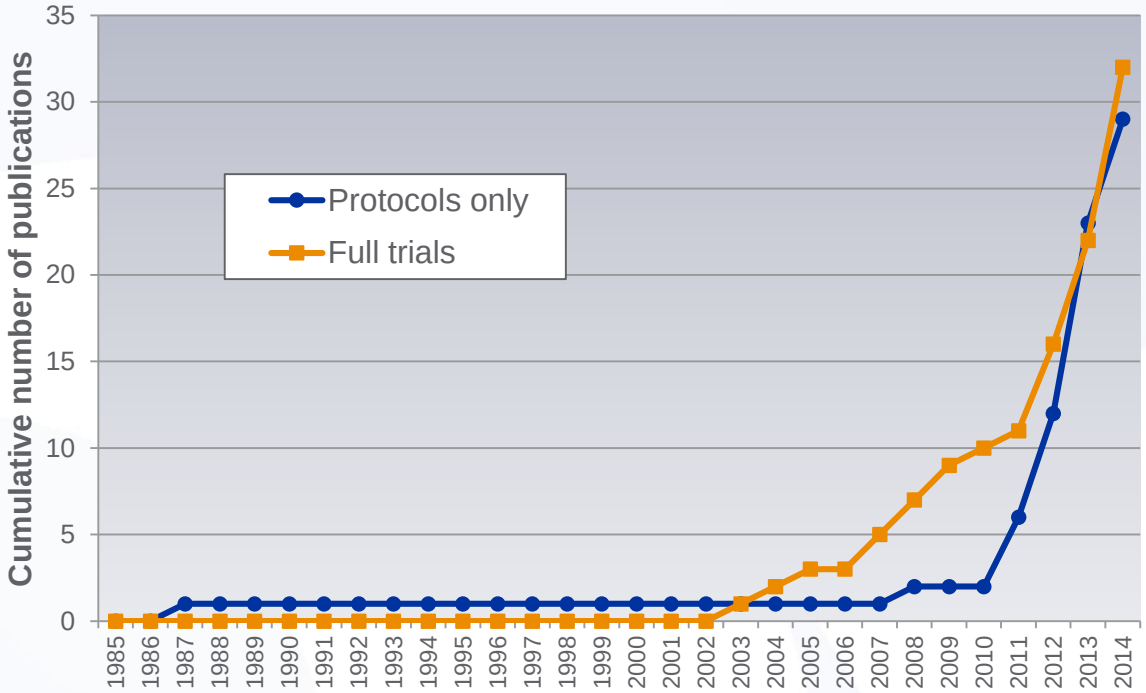
DESIGN OPTION 3: LONGITUDINAL PARALLEL ARM CRT



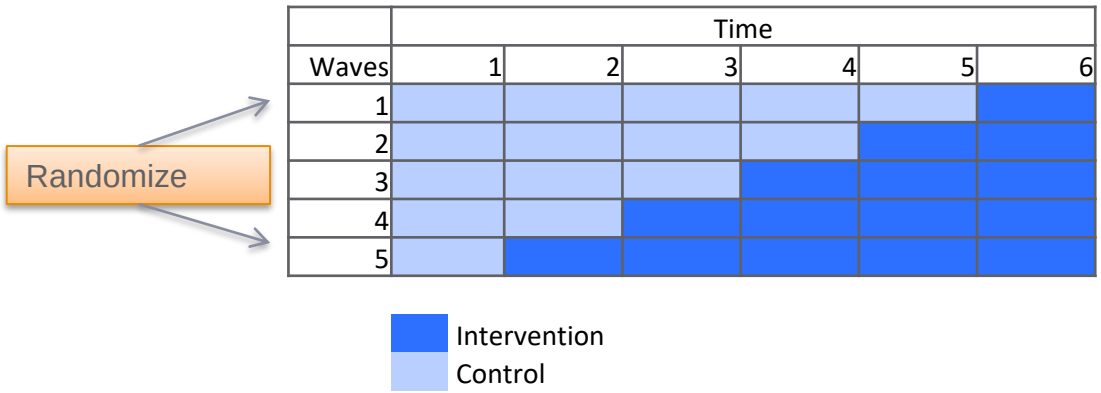
- Disadvantages:
  - Takes longer to complete the study
  - May increase the risk of hospitals dropping out
  - May increase the risk of contamination (e.g., control hospitals adopting the intervention on their own) or external events influencing outcomes
  - More complicated to analyze

A NOVEL DESIGN: THE STEPPED WEDGE

- Published stepped wedge trials, 1985-2014



DESIGN OPTION 4: THE STEPPED WEDGE CRT



- All hospitals start in control and end in intervention condition
- Hospitals cross to intervention sequentially and in random order
- Outcomes are assessed repeatedly in each hospital over time
- Intervention effect is a combination of within-site and between-site differences

ADVANTAGES OF THE STEPPED WEDGE

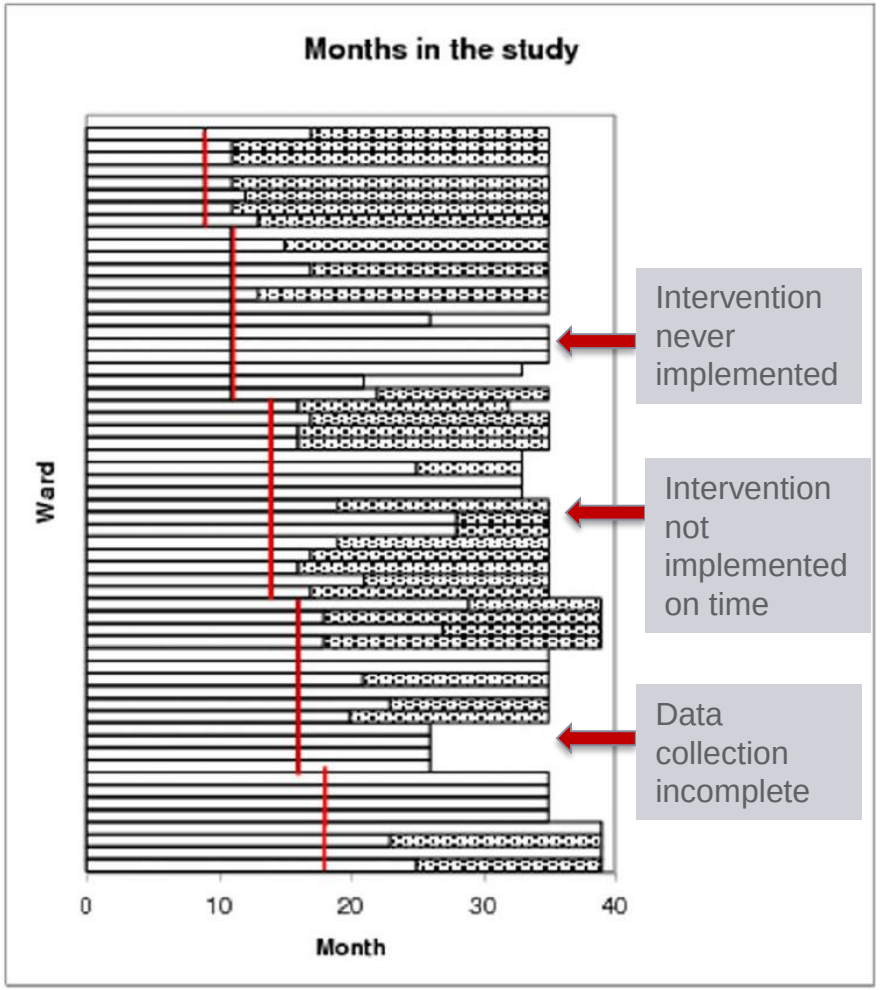
- All hospitals receive the intervention during the study
  - Easier to recruit hospitals if they know they will receive the intervention
  - Stakeholders may *require* that the intervention be implemented across the entire health system in order to exert its anticipated benefits
- Increases “power” (ability to detect a difference) over parallel arm designs
- Delivery of the intervention can be spread out over time (e.g., by having only one hospital cross over each time)

DISADVANTAGES OF THE STEPPED WEDGE

1. All participating hospitals must be recruited upfront (so they can be randomized)
2. Can be logistically challenging to ensure all sites are ready to implement on schedule

WHEN THINGS GO WRONG

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



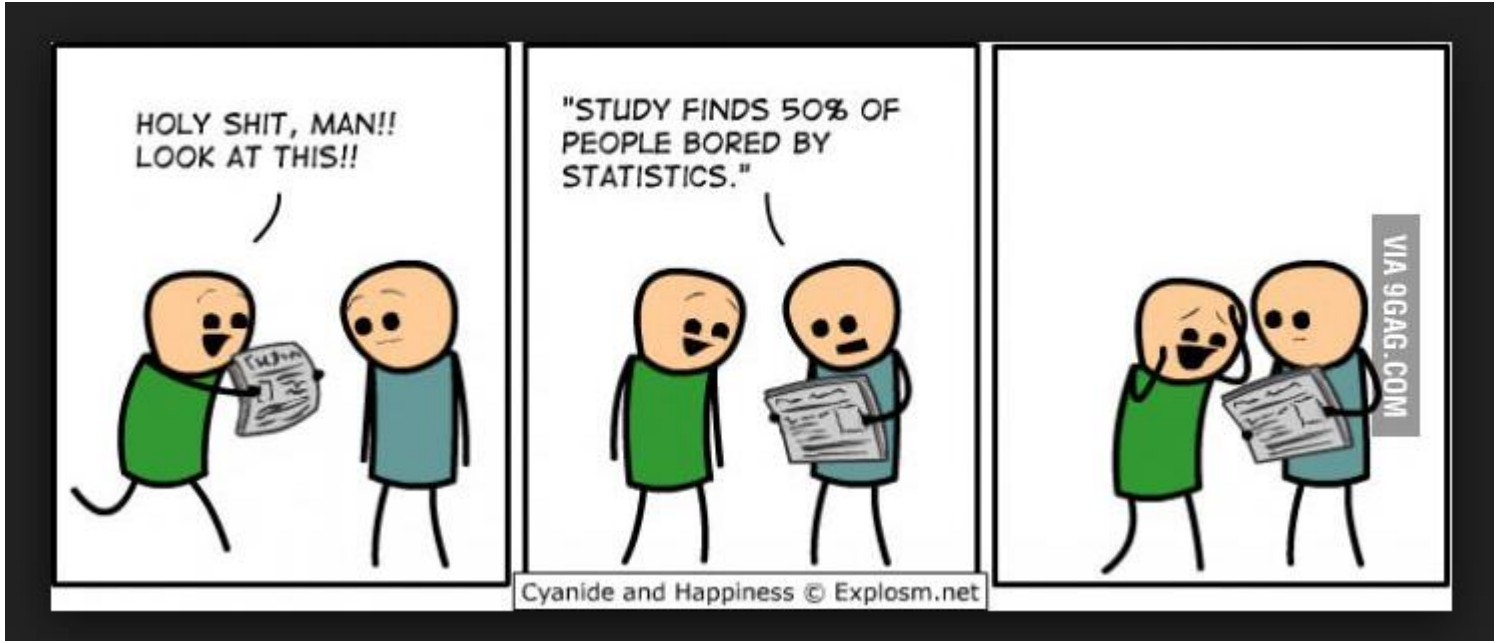
DISADVANTAGES OF THE STEPPED WEDGE

1. All participating hospitals must be recruited upfront (so they can be randomized)
2. Can be logistically challenging to ensure all sites are ready to implement on schedule
3. Takes longer to complete the study
4. May increase the risk of hospitals dropping out
5. May increase the risk of contamination
6. More complicated to analyze and interpret results (requires many assumptions)
7. Does not work well if intervention does not have an immediate effect
8. Does not work well if intervention effect might change over time
9. Can be difficult to separate the effect of the intervention from the effect of secular trends

EXAMPLE: THE PRE-OP TESTING TRIAL

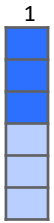
- **Key question for the statistician:** *How many hospitals should we include in the study?*
  - Trade-off between having a big enough sample to detect an effect of the intervention while keeping the study affordable
  - Hoping to detect a 7% drop in the use of unnecessary tests
- Different study designs have different implications for the required number of hospitals
- *Which study design minimizes the sample size?*

…NOW DOING FANCY STATISTICS…



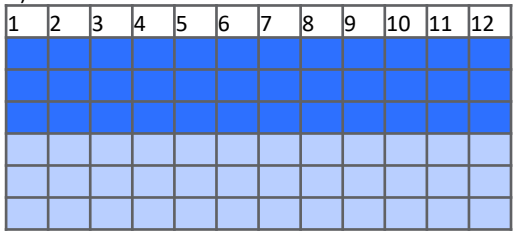
REQUIRED # HOSPITALS: PARALLEL ARM DESIGNS

1) Parallel arm CRT over 1 month



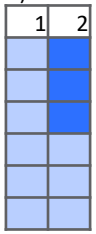
292 hospitals
58,400 surgeries

1) Parallel arm CRT over 12 months



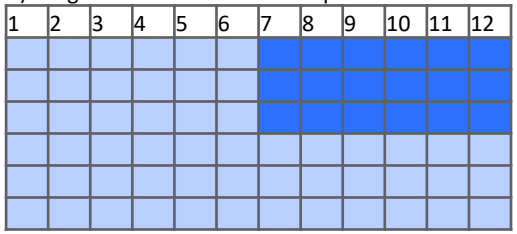
234 hospitals
561,600 surgeries

2) Before and after parallel arm CRT over 2 months



116 hospitals
46,400 surgeries

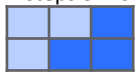
3) Longitudinal before and after parallel arm CRT



22 hospitals
52,800 surgeries

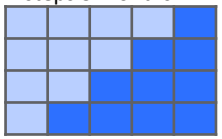
REQUIRED # HOSPITALS: STEPPED WEDGE DESIGNS

2 steps 3 months



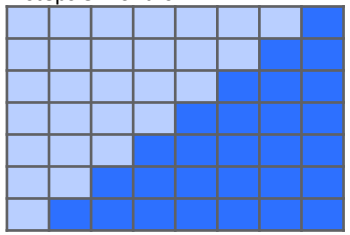
93 hospitals
55,800 surgeries

4 steps 5 months



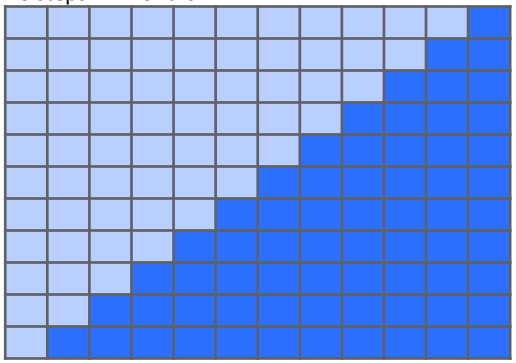
42 hospitals
42,000 surgeries

7 steps 8 months



25 hospitals
40,000 surgeries

10 steps 11 months



16 hospitals
38,400 surgeries

CONCLUSIONS

- Cluster randomized trials are the gold standard study design for evaluating implementation strategies
- Traditionally, we randomize sites to either intervention and control arms and observe outcomes at the end of the trial (parallel arm CRTs)
- More powerful (cost-efficient) designs include multiple periods of observations before and after intervention
- The stepped wedge is a novel longitudinal design which evaluates roll-out of an intervention across an entire health system
- Advantages and disadvantages of the stepped wedge design need to be carefully considered before adopting this design

DESIGNS WITH MULTIPLE BEFORE AND AFTER PERIODS

- Multiple periods before and after the rollout of the intervention don't contribute useful information ("inefficient")…

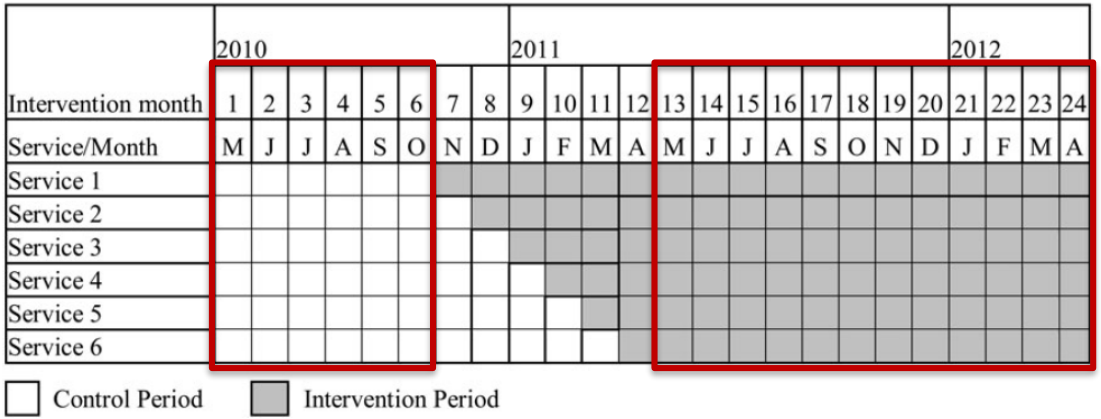


Figure 1. Timeline of antimicrobial stewardship stepped-wedge rollout to the clinical services. The "stepped" introduction of the intervention occurred over months 7–12; the white "wedge" reflects the control period, and the gray "wedge" the intervention period.

Clinical Infectious Diseases 2014;59(6):867–74

DESIGNS WITH MULTIPLE BEFORE AND AFTER PERIODS

- … But allows approaching the design as a “multiple baseline design”

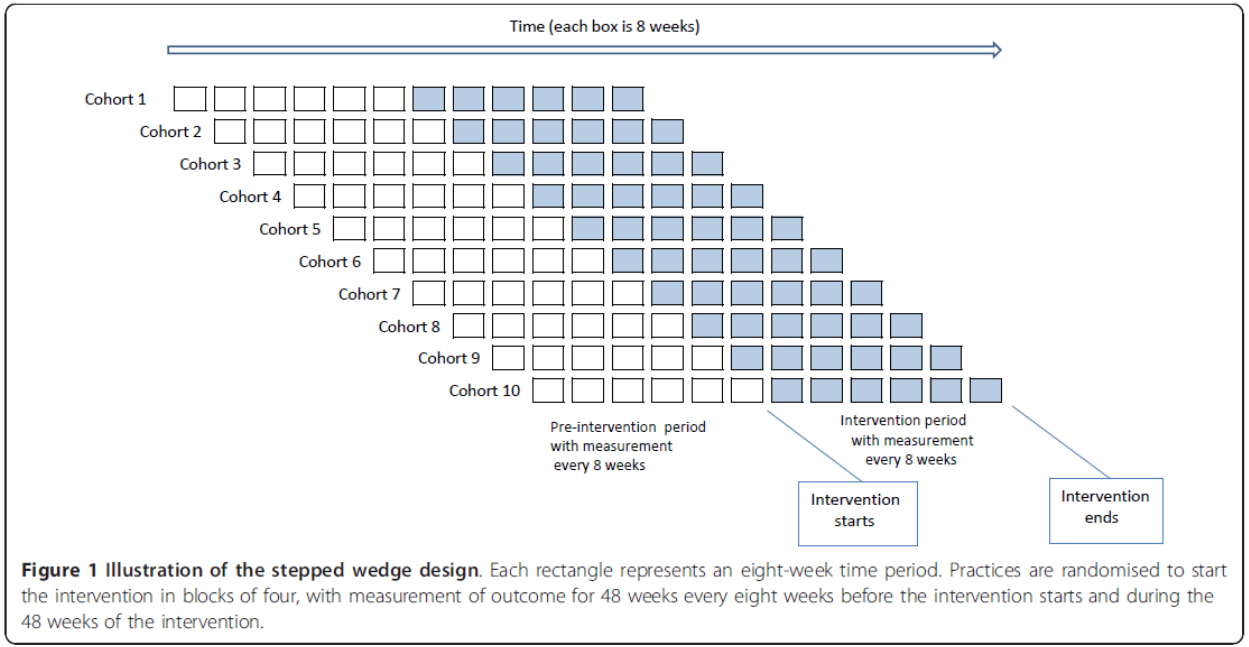
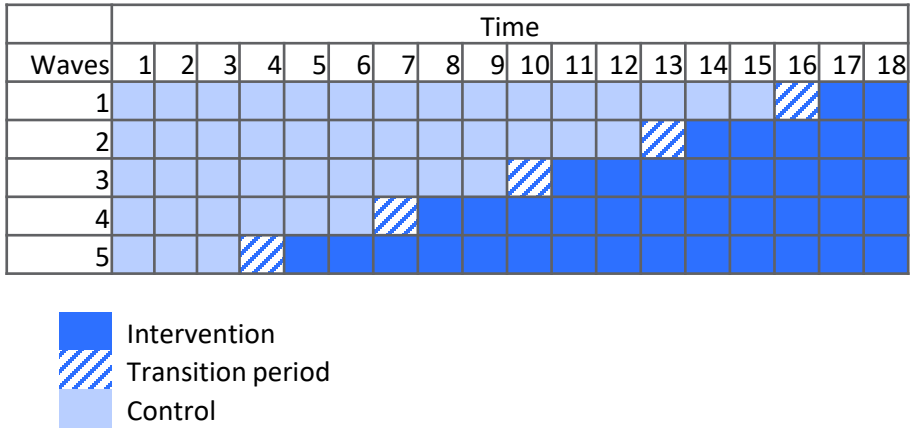


Figure 1 Illustration of the stepped wedge design. Each rectangle represents an eight-week time period. Practices are randomised to start the intervention in blocks of four, with measurement of outcome for 48 weeks every eight weeks before the intervention starts and during the 48 weeks of the intervention.

Dreischulte *et al. Implementation Science* 2012, **7**:24
http://www.implementationscience.com/content/7/1/24

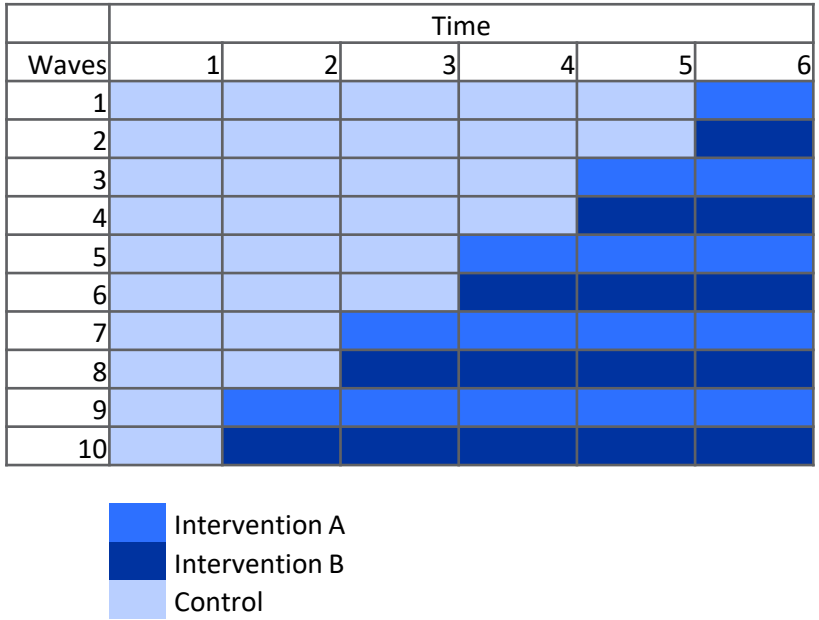
STEPPED WEDGE WITH TRANSITION PERIOD

- Can allow for a short transition period to allow the intervention to be put in place



MULTI-ARM STEPPED WEDGE (1)

- Head to head comparison of two interventions



MULTI-ARM STEPPED WEDGE (2)

- Testing a sequential (add-on) intervention