STEPPED WEDGE DESIGN IN PRACTICE: A CHECKLIST FOR FEASIBILITY

Diane L. Fairclough, DrPH
Erin Leister Chaussee, PhD

ACCORDS
ADULT AND CHILD CONSORTIUM FOR HEALTH OUTCOMES RESEARCH AND DELIVERY SCIENCE
UNIVERSITY OF COLORADO | CHILDREN'S HOSPITAL COLORADO

Outline

- SW design overview
- Checklist for feasibility
- Example: DECIDE-LVAD
- Other challenges

The stepped wedge design

- Quasi-experimental design
  - Hybrid of cluster randomized and cross-over
  - Crossover is unidirectional (C => I)
  - Time of crossover is randomized
- Two versions
  - Cross sectional – enrollment of individuals is continuous, time of enrollment determines treatment assignment
  - Cohort – individuals enrolled at beginning; crossover from C to I occurs within individual

	Time			
1	2	3	4	5
C	I	I	I	I
C	C	I	I	I
C	C	C	I	I
C	C	C	C	I

Reasons for choosing the SW Design

- Evaluate the “effectiveness” or the implementation of an intervention previously shown to be efficacious in an individually randomized trial or in a different setting; systematically evaluate new program
  - **Effectiveness** – all sites participate in the intervention, can continue past end of study
  - **Implementation** – able to study the implementation more carefully as that is spread out over time

Reasons for choosing the SW Design

- **Efficiency**: Units act as their own control (same as cross-over design) => Smaller sample size than cluster randomized design when ICC is large (will define later)
- **Logistical or financial** - cannot introduce the intervention in all units at once; need to study implementation
- **Recruitment of sites** (more willing to participate as all will participate in the intervention)

Statistical Model

Model:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + X_{ij}\theta + e_{ijk}$$

Hussey & Hughes, *Contemp Clin Trials* 2007

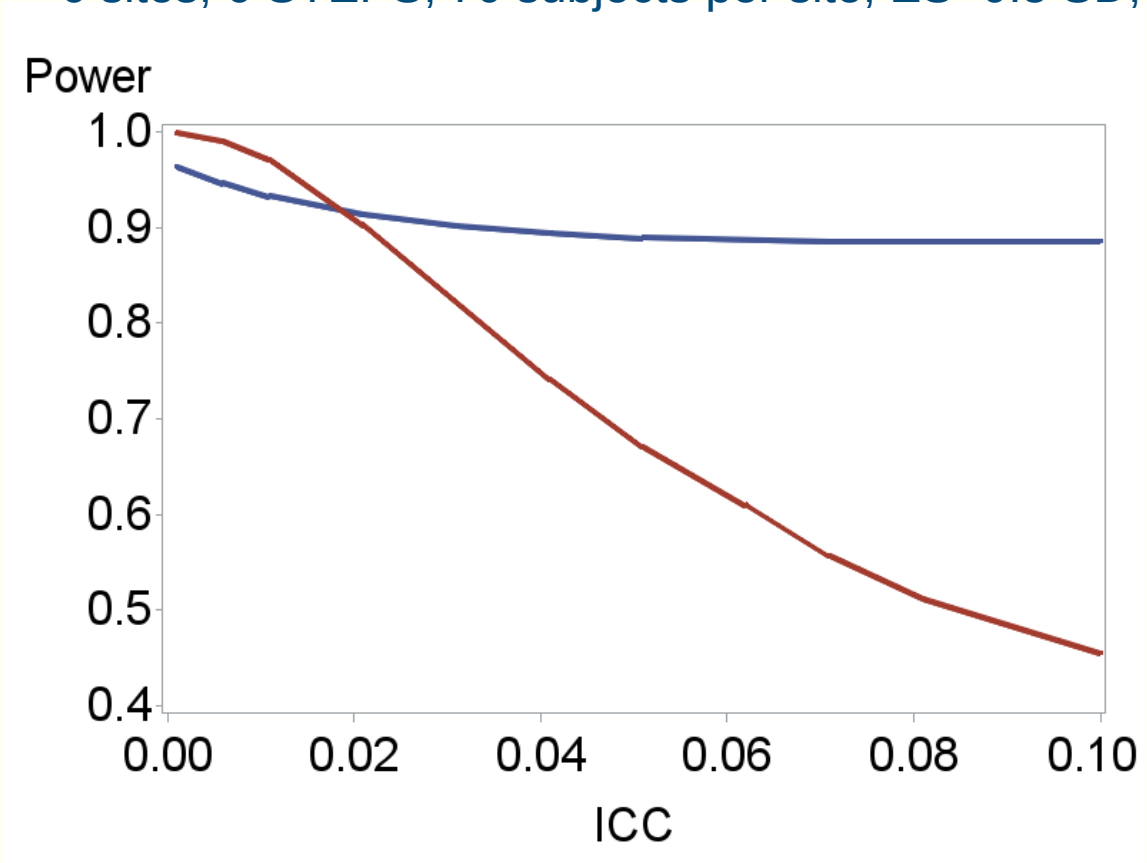
θ	treatment effect
β_j	time effects (constant across cluster)
$\alpha_i \sim N(0,\tau^2)$	between cluster variation
$e_{ijk} \sim N(0,\sigma^2)$	within cluster variation

Key issue determining the power/sample size in a CRT:

$$\text{ICC: } \text{Corr}(Y_{ijk}, Y_{ij'k'}) = \tau^2/(\tau^2 + \sigma^2) \neq 0$$

Power – SW vs CRT

6 sites, 6 STEPS, 70 subjects per site, ES=0.5 SD, alpha=0.05



stepped wedge

cluster randomized

Best thing since sliced bread?

- Why would we not always want to use a SW design vs. a Cluster Randomized Trial?

- Experience with a SW trial (DECIDE) and work on other proposals

Checklist for feasibility

Checklist part 1: Administrative

- ❑ Is it feasible to start enrollment at all the sites at the same time?
  - Coordinating IRBs and subcontracts
  - Hiring of study personnel at each site

- ❑ Are all the sites likely to complete the study (e.g. site dropout is unlikely)
  - Commitment of all sites to complete the study

Part 2: Accrual and Selection Bias

- ❑ Are all sites committed to similar levels of accrual during both control and intervention phases of the trial?
- ❑ Is the pool of potential participants large enough (or continually renewing) to avoid biased selection over time?
  - Very large or renewing pool of participants
  - Steady recruitment that is consistent with respect to patient characteristics (no selection bias)
    - Avoid selection based on visits (sicker patients have more visits)
  - Avoid change in eligible subjects due to intervention

Part 3: Cross-over

- Can the cross-over occur at a specific point in time?
  - Is the **duration of the intervention** short enough to avoid contamination during the cross-over phase of the trial or contamination unlikely?
  - Is the **duration of follow-up** of participants short enough to avoid contamination or is contamination unlikely during follow-up?
  - How much **training** is required? Does the intervention require practice to deliver effectively?

Red Flags/Need for Washout

- Crossover (XO) requires training/practice to implement intervention
- Extended duration of the intervention
  - What do you do with subjects recruited just before the XO
- Follow-up: Potential for contamination of FU after the XO

Washout/Rollout

- Avoiding contamination between the control and intervention phases: extended interventions, extended follow-up
- Training to achieve full effect of intervention

		Time Period						
Treatment sequence	**Site**	**1**	**2**	**3**	**4**	**5**	**6**	
1	**1 Site**	C	WR	I	I	I	I	**Key**
2	**2 Sites**	C	C	WR	I	I	I	C **Control**
3	**2 Sites**	C	C	C	WR	I	I	WR **Washout/Rollout**
4	**1 Site**	C	C	C	C	WR	I	I **Intervention**

N.B. Incomplete designs will dramatically impact the power

Impact of WR on Power

Complete design

	Time						
#	1	2	3	4	5	6	7
1	C	I	I	I	I	I	I
2	C	C	I	I	I	I	I
3	C	C	C	I	I	I	I
4	C	C	C	C	I	I	I
5	C	C	C	C	C	I	I
6	C	C	C	C	C	C	I

Washout/rollout (WR)

	Time							
#	1	2	3	4	5	6	7	8
1	C	WR	I	I	I	I	I	I
2	C	C	WR	I	I	I	I	I
3	C	C	C	WR	I	I	I	I
4	C	C	C	C	WR	I	I	I
5	C	C	C	C	C	WR	I	I
6	C	C	C	C	C	C	WR	I

Scenario	Total N	C or I N/cell	WR N/cell	*Power
Complete	840	20	NA	0.80
WR	840	20	0	0.66
WR	1176	28	0	0.79

*6 sites, 7\8 time periods (complete\WR), alpha=0.05, ICC=0.1

Part 4: External/Internal Time Trends

- Is it unlikely that events or changes in policy might result in time trends that occur in some sites but not all sites?
  - Maybe alternative analytic models, but should be pre-specified. Cutting edge with respect to analysis (e.g. evolving literature).

Example: DECIDE-LVAD

- Drs Larry Allen and Dan Matlock, UC-SOM
- **Population** – Patients with heart failure considering LVAD implantation (and caregivers)
- **Intervention** – Decision aid
- **Goals:** Study Effectiveness and Implementation
- **Outcomes**
  - LVAD knowledge (~1 week FU),
  - Values-treatment concordance (1 month FU)
- **SW details** – 6 LVAD centers across the US (CU required to start first)

Example: DECIDE

	Enrollment/Time Period					Total	
Site	1	2	3	4	5	Cntrl	Invtn
1	13	8	12	7	1	13	28
2	5	6	4	3	4	11	11
3	4	9	16	15	11	13	42
4	10	7	8	7	4	25	11
5	8	8	12	5	2	28	7
6	18	11	6	10	14	45	14
Total	58	49	58	47	36	135 (54%)	113 (46%)
Outpatient (%)	16%	14%	29%	30%	31%	17%	31%
HF Dx > 4 yrs (%)	82%	84%	65%	75%	68%	77%	64%

DECIDE: Checklist Administrative

- ❑ Is it feasible to start enrollment at all the sites at the same time?
  - One site started late (only by one month) – due to IRB/subcontract administrative approval process

- ❑ Are all the sites likely to complete the study (e.g. site dropout is unlikely)
  - One site stopped enrollment for a short time during the middle of the study period

Ref: Matlock et al

DECIDE: Checklist Accrual

- ❑ Are all sites committed to similar levels of accrual during both control and intervention phases of the trial?
  - Accrual declined in most sites over time, except one site
  - Control: N=135, Intervention: N=113
  - Original Target: 65 per arm.
- ❑ Is the pool of potential participants large enough (or continually renewing) to avoid biased selection over time?
  - Proportion of Outpatients increased over time
    → 17% in Control; 31% in Intervention
  - Proportion with Initial DX >4 years decreased over time
    → 77% in Control; 64% in Intervention

DECIDE: Checklist Cross-over

- ❑ Can the cross-over occur at a specific point in time?
  - ❑ Is the **duration of the intervention** short enough to avoid contamination during the cross-over phase of the trial or contamination unlikely?
    - Short duration of intervention for participants – viewing of pamphlet and video
  - ❑ Is the **duration of follow-up** of participants short enough to avoid contamination or is contamination unlikely during follow-up?
    - Contamination unlikely during follow-up – outcomes measured at ~1 week and 1 month
  - ❑ How much **training** is required? Does the intervention require practice to deliver effectively?
    - Site training: study staff visited each site right before cross-over. No practice needed, but changes such that pamphlet and video are offered

DECIDE: Checklist Time Trends

- ❑ Is it unlikely that events or changes in policy might result in time trends that occur in some sites but not all sites?
  - This was considered unlikely – potential events (celebrity implantation of LVAD, standard of care/treatment/industry materials) were considered to be likely to change things across all sites, if they occurred

Other Challenges

- Sample size calculations:
  - Not difficult if all the sites are the same size and can recruit the same number of participants during each period
  - More challenging if
    - site sizes are variable,
    - accrual over time is variable,
    - washout is needed at crossover
    - other non-standard SW design or hybrid

Other Challenges (2)

- Randomization
  - Balancing sites by size
    - Identifying large and small sites
  - Balancing site level covariates
  - Balancing participant level covariates

Other Challenges (3)

- Analysis methods
  - Hussey and Hughes model is straight forward and quite robust if confounding covariates are measured and included in the model, but may not address all concerns
  - Varying time effects by cluster or groups of clusters
    - See Hooper et al (2016), Hemming et al (2017)
  - Varying treatment effect by cluster or groups of clusters
    - See Hemming et al (2017), Thompson et al (2017)
  - Treatment effect varying over time
    - See Hemming et al (2017)
  - Other robust methods:
    - Ji et al (2017), Wang & De Gruttola (2017), Thompson et al (2018), Hughes et al (2019)

Summary – To SW or Not?

- Checklist
- Strong statistical support
- Other discussions
  - deHoop (2015) BMC Med Res Method
  - Hargreaves (2015) Trials
  - Hemming (2015) BMJ
  - Taljaard (2016) Clinical Trials
  - Eichner (2019) JCE

References

1) Hussey MA, Hughes JP. Design and analysis of stepped wedge cluster randomized trials. Contemporary Clinical Trials 2007

2) Hemming et al. The stepped wedge cluster randomized trial: rationale, design, analysis, and reporting. BMJ 2015

3) Taljaard et al. Substantial risks associated with few clusters in cluster randomized and stepped wedge designs. Clinical Trials 2016

4) De Hoop et al. The need to balance merits and limitations from different disciplines when considering the stepped wedge cluster randomized trial design. BMC Med Res Method 2015.

5) Hargreaves et al. Five questions to consider before conducting a stepped wedge trial. Trials 2015.

6) Hemming et al. Analysis of cluster randomized stepped wedge trials with repeated cross-sectional samples. Trials 2017.

7) Girling and Hemming. Statistical efficiency and optimal design for stepped cluster studies under linear mixed effects models. Statistics in Medicine 2016.

8) Eichner et al. Systematic review showed that stepped-wedge cluster randomized trials often did not reach their planned sample size. J Clin Epi 2019.

References (2)

9) Hooper et al (2016) Sample size calculation for stepped wedge and other longitudinal cluster randomised trials. Statistics in Medicine

10) Hemming et al (2017) Analysis of cluster randomised stepped wedge trials with repeated cross-sectional samples. Trials

11) Thompson et al (2017) Bias and inference from misspecified mixed-effect models in stepped wedge trial analysis. Statistics in Medicine

12) Ji et al (2017) Randomization inference for stepped-wedge cluster-randomized trials: An application to community-based health insurance. Ann Appl Stat

13) Wang & De Gruttola (2017) The use of permutation tests for the analysis of parallel and stepped-wedge cluster-randomized trials. Statistics in Medicine

14) Thompson et al (2018) Robust analysis of stepped wedge trials using cluster-level summaries within periods. Statistics in Medicine

15) Hughes et al (2019) Robust inference for the stepped wedge design. Biometrics

References (3)

DECIDE-LVAD References

- McIvennan et al. A Multicenter Trial of a Shared Decision Support Intervention for Patients and Their Caregivers Offered Destination Therapy for Advanced Heart Failure: DECIDE-LVAD Rationale, Design, and Pilot Data. J Cardiovasc Nurs 2016.

- Allen et al. Effectiveness of an Intervention Supporting Shared Decision Making for Destination Therapy Left Ventricular Assist Device: The DECIDE-LVAD Randomized Clinical Trial. JAMA IM 2018.

- Matlock et al. Decision Aid Implementation among Left Ventricular Assist Device Programs Participating in the DECIDE-LVAD Stepped-Wedge Trial. Medical Decision Making 2020.