Certificate Program in Practice-Based Research Methods

PBRN Methods: Clustered Designs

Session 8 - January 26, 2017

L. Miriam Dickinson, PhD
Professor, University of Colorado School of Medicine
Department of Family Medicine

Donald E. Nease, Jr., MD
Associate Professor
Vice Chair of Research
Director of Community Engagement & Research, Colorado Clinical and Translational Sciences Institute
Director, Shared Networks of Collaborating Ambulatory Practices & Partners (SNOCAP)

PBRN Methods: Clustered Designs

L. Miriam Dickinson, PhD
Donald Nease, MD
University of Colorado

Content

- Background: Review some common research approaches and study designs
- Clustering: a common Features of PBRN research
- Cluster Randomized Trials
- Sample size and power
- Randomization
- Data analysis
- Heterogeneity of treatment effects
- Stepped wedge designs

Background: Review of Some Common Concepts and Study Designs

- **Retrospective**: start with present and look backward at subject's history (e.g. case-control)
- **Cross-sectional**: a snap-shot of subjects at one point in time
- **Prospective**: start with present and follow subjects into the future (cohort study)
  - Retrospective cohort
- Person-time studies – time to event
- Traditional randomized controlled trial (RCT)
- Cluster randomized trial
- Stepped Wedge Trials
- Related concepts:
  - The RE-AIM Framework: Reach, Effectiveness, Adoption, Implementation, Maintenance (Glasgow)
  - Comparative effectiveness and pragmatic trials
  - Implementation and dissemination

Clustering: a common feature of PBRN Research

- Study designs can be observational or experimental
- Retrospective, cross-sectional, or prospective time frames can all be used
- Clustering, or nesting, is a common feature of PBRN research and can apply to any of the common study designs
  - Primary type of PBRN clustering usually involves patients nested within practices (sometimes patients within clinicians within practices)
  - Can include repeated observations on patients over time (longitudinal studies) as well
  - Many studies in PBRNs have both kinds of clustering
- Study design, sampling approaches, power, statistical analysis are *all* affected by clustering

Clustered Randomized Trial (CRT)

- CRTs are a variant of the traditional randomized controlled trial
- Randomized controlled trial is classic experimental study
  - *Patients* are randomly assigned to one of two or more groups (e.g. usual care or intervention) and we observe them to see if the intervention improves outcomes
- In clustered randomized trials in PBRNs the unit of randomization is generally the practice (occasionally some geographic unit, such as communities or counties)
- Some typical reasons for cluster instead of patient level randomization
  - Interventions may target the practice/environment rather than the patient per se
  - Contamination
  - Logistical, cost, and/or ethical concerns
- CONSORT statement: see extension for CRTs
  - http://www.consort-statement.org

Designing a CRT: an example

- Start with your research question
  - Design and analysis should directly address research question and be congruent with the conceptual model
- Example
  - Research questions:
    - Will a practice facilitation approach based on the chronic care model improve patient care and clinical outcomes for diabetic patients
  - Rationale for choice of study design
    - Implementing the intervention within a practice will likely affect all patients, thus contamination would be a serious problem for a traditional RCT
  - CRT hypotheses will be a little different than in a traditional RCT
    - Improvement in quality of diabetes care will be greater for patients in practices receiving the intervention than patients in usual care
    - Improvement in HbA1c will be better for patients in practices receiving the intervention than patients in usual care

Sample Size

- How many practices? How many patients?
  - Involve a biostatistician early in the planning stage and throughout the study
- Power analyses based on number of patients have to be adjusted for clustering
- Intraclass correlation coefficient (ICC) measures the similarity of patients within practices compared to patients in other practices
  - Proportion of the total variance in outcome variable(s) accounted for by clustering – often expressed as a %
- Example: For the primary outcome of HgA1c, previous work indicates that the ICC is about 5%, ICCs for process of care outcomes can be much higher, often as high as 10%

- References : Donner A, Klar N. *Design and Analysis of Cluster Randomization Trials in Health Research*. London, England: Oxford University Press; 2000.
- Dickinson LM, Basu A. Multilevel modeling and practice-based research. Ann Fam Med. 2005 May-Jun;3 Suppl 1:S52-60.

How to determine sample sizes for CRTs

- Determine your primary outcome variables
- Obtain an estimate of the ICC (actual data, literature, this can be challenging, sometimes we just have to guess)
- Calculate the variance inflation factor (VIF): (1 + (m – 1)ICC), where m is the number of patients per practice
- Calculate the effective sample size: divide the proposed sample size (m x number of practices) by the VIF to get the effective sample size
- Do a traditional power analysis

Practices per group	Patients per practice	ICC	VIF	Effective sample size	Effect size	power
6	50	5%	3.45	87	.43	>80%
6	50	10%	5.9	51	.56	80%
6	50	15%	8.35	36	.67	80%
6	100	10%	10.9	55	.55	>80%
10	50	10%	5.9	85	.44	>80%

Randomization

- Now that we know how many practices/patients we need, how do we assign them to groups?
  - Often we recruit just enough practices to do the study
  - Occasionally we have the luxury of sampling practices from a larger pool: stratified sampling may help in this case
- Generally, the number of practices to be randomized is much smaller than trials in which individuals are randomized
- Heterogeneity among practices
- Individuals within practices are more similar to each other than members of other practices
- Simple randomization can result in study arms that are very different from each other, resulting in covariate imbalance between study arms
- Stratified randomization can improve balance but doesn't always solve the imbalance problem
- Minimization methods extended to CRTs

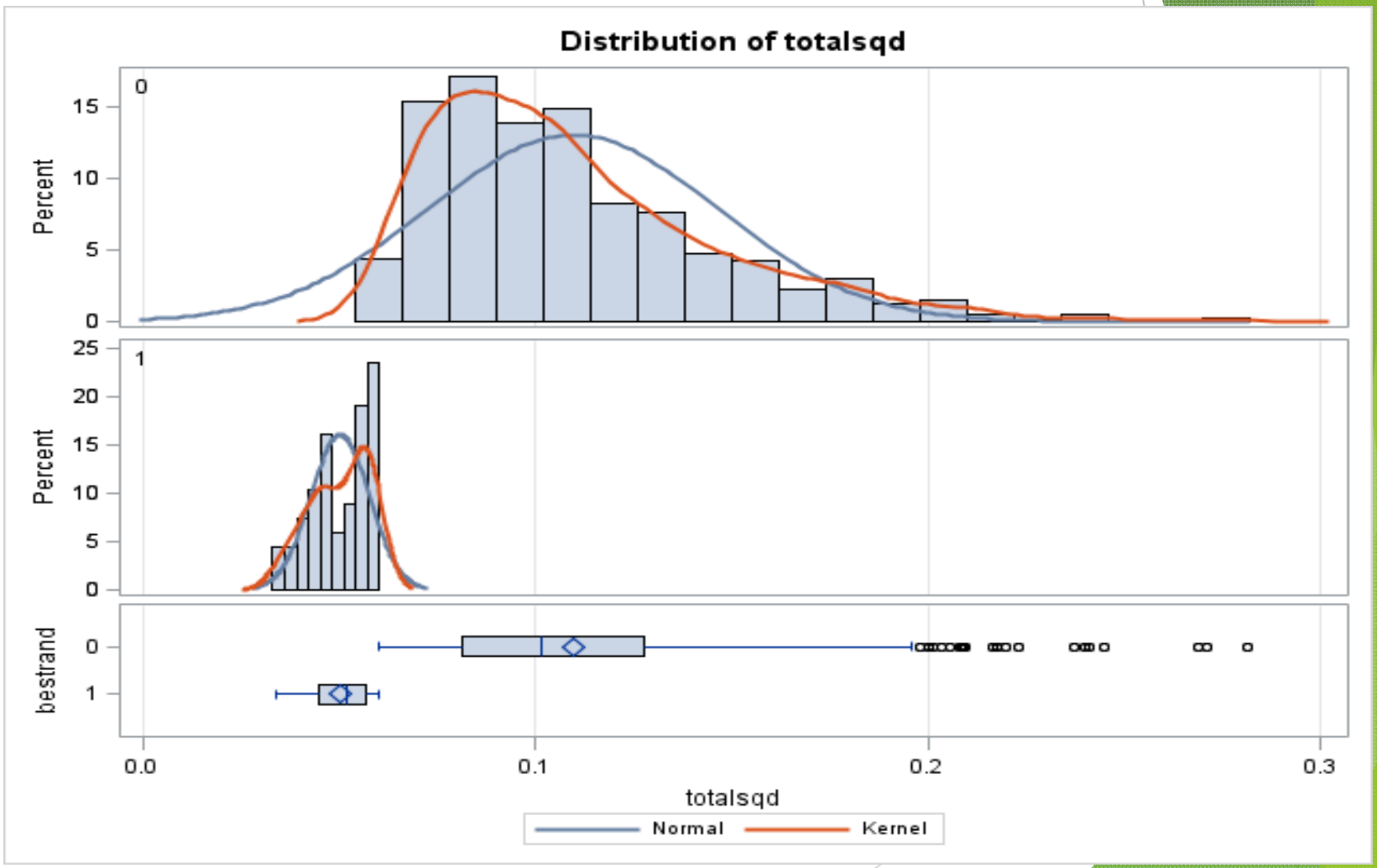
Covariate Constrained Randomization

- Particularly useful for PBRNs is baseline data are available (usually summary data from practices)
- All possible randomizations of units into study groups are generated
- A balance criterion (B), defined as the sum of squared differences between study groups on relevant standardized variables, is calculated for each randomization
  - $B=(w_1(x_{11} - x_{21})^2 + w_2(x_{12} - x_{22})^2 + \ \ldots\ )$
  - Where w is the weight for each selected variable, x11 is the mean for group 1, variable 1, x21 is the mean for group 2, variable 1, etc.
- Establish a criterion for maximum allowable difference between study groups and define a set of "acceptable randomizations" in which the differences between treatment groups on covariates are minimized
- A single randomization is then chosen from the set of "acceptable randomizations"
  - See Dickinson, et al, Pragmatic cluster randomized trials using covariate constrained randomization: A method for practice-based research networks (PBRNs). JABFM. 2015 Sep-Oct;28(5):663-72.

Covariate Constrained Randomization Example: CKD Study Description

- Study objective: To test two approaches to improving care for stage 3 and 4 CKD patients in primary care practices based on the Chronic Care Model (CCM)
- Variables for Randomization aggregated to the practice level
  - Structural and sociodemographic data
    - # FTE clinicians, % African American, % Hispanic, % Medicaid or uninsured
  - Clinical data
    - % of patients with HbA1c>9, % diabetic, % stage 4 CKD, % with systolic BP>130, % with systolic BP>140
    - Mean GFR, mean HbA1c, mean systolic BP
  - Stratification variables handled as part of the procedure by restricting to randomizations with a pre-specified number in each arm by identified strata
  - Achieved balanced study arms (i.e. no significant baseline differences between study arms on aggregated practice level variables)

Distribution of balance criterion



Data Analysis for CRTs

- Describe the sample and address issues of external and internal validity
  - Clustering adds a level to be considered in the CONSORT diagram
    - Describe retention at both the practice and patient level
  - How representative are the patients and practices in this study of the target population? (external validity)
    - Describe practice characteristics
    - Describe patient characteristics
  - Did the randomization work?
    - Compare patients in the intervention group to controls on key variables
    - Variables that differ significantly between groups should be included as covariates in analyses
  - Analytic approaches for clustered data
  - Missingness
  - Did the intervention work?

Did the randomization work?

- Are control and intervention groups similar on key baseline characteristics?
- Compare control and intervention groups on baseline practice and patient characteristics using t-tests and chi-square tests (unadjusted and/or adjusted)
- Example: CRT of two practice facilitation approaches to standard care for improving patient care and clinical outcomes for diabetic patients
  - Stratified randomization approach
  - Practices were similar with regard to rural vs urban location and % Medicaid
  - Patients were similar in terms of sociodemographic characteristics but differed somewhat on clinical variables
  - Baseline process of diabetes care (POC) differed between study arms
    - POC: sum of nine items from the American Diabetes Association Physician Recognition Program: HgA1c, foot exam, blood pressure, dilated eye exam, cholesterol, nephropathy screen, flu shot, nutrition counseling, and self-management support
- See Dickinson WP, et al, Practice Facilitation to Improve Diabetes Care in Primary Care: A Report from the EPIC Randomized Clinical Trial Annals of Family Medicine. 2014; 12(1)8-16.

Longitudinal studies: very important to assess for mechanisms of missingness

- Simplest form of missingness is patient dropout sometime after baseline
  - For dropouts vs completers
    - Compare baseline characteristics using chi-square tests, t-tests, Kendall's tau
    - Compare values of the outcome variable at all observed timepoints
- Key terms
  - MCAR: Missing completely at random – missingness not associated with any observed variables
  - MAR: Missing at random – missingness associated with baseline or subsequent observed variables
  - MNAR: Missing not at random – missingness associated with unobserved characteristics (i.e. patient becomes very ill and drops out)
  - MCAR and MAR are "ignorable" and can be handled analytically using likelihood based models with covariates associated with missingness included
  - MCAR is "non-ignorable" and requires special approaches
- See Fairclough book for more complex situations, including non-ignorable missingness
  - Fairclough DL: Design and analysis of quality of life studies in clinical trials. New York, Chapman & Hall/CRC, 2010

Data Analysis for CRTs

- Some common analytic methods for non-clustered data
  - Simple stats for associations: chi-square tests and t-tests (we use these to compare study groups on baseline data)
  - Multiple logistic (dichotomous outcome) or linear (continuous outcome) with categorical or continuous predictors
  - Survival analysis (e.g. Cox proportional hazards) – outcome is time to event
- Some common analytic approaches for clustered data
  - General (or generalized) linear mixed models (GLMM) (sometimes called mixed effects regression models, multilevel or hierarchical models)
    - Often used to adjust for clustering (e.g. patients within practices) or longitudinal studies with repeated measures on patients, or both
    - GLMMs are a likelihood based approach and can accommodate certain kinds of missing data as well as clustering
  - Generalized estimating equations (GEE)
    - Sometimes used instead of GLMMs, assumptions are different
  - Survival analysis (Cox proportional hazards) - clustered survival analysis can be done

Did the intervention work?

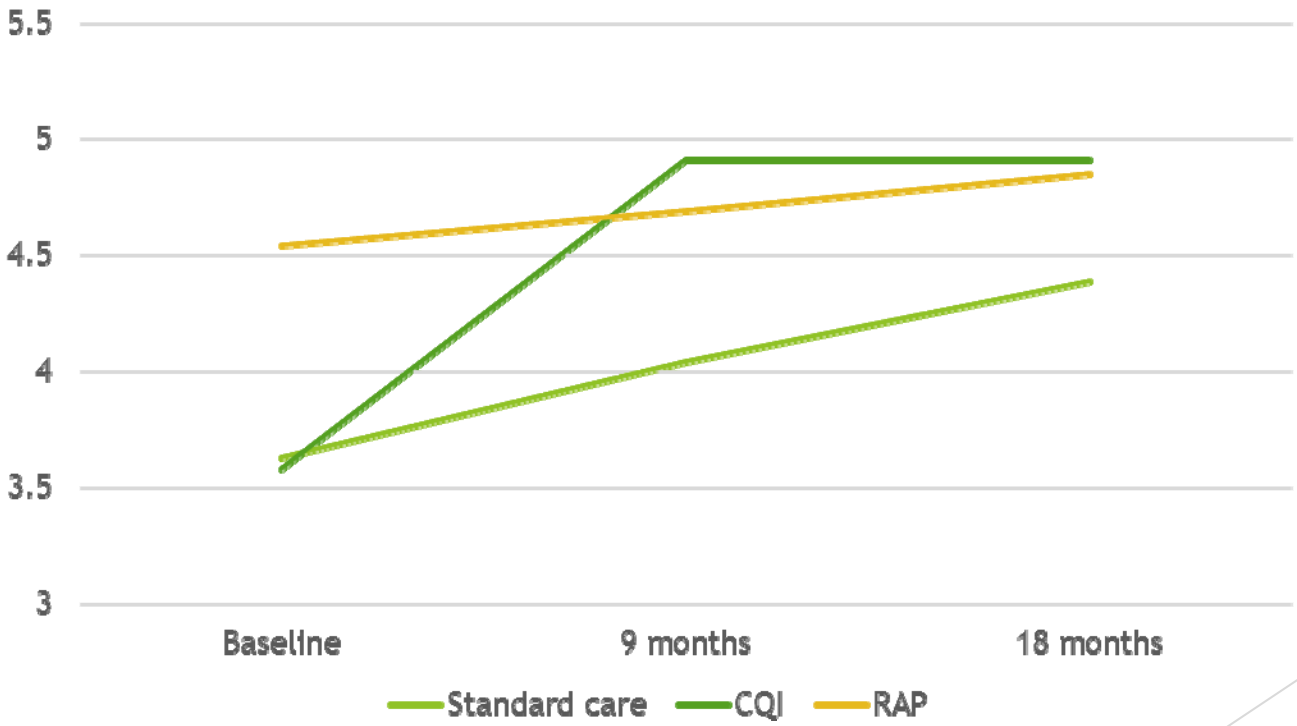
- Since the data are clustered, general linear mixed effects models used for analysis
- Random effect for patient and practice
- Outcome: diabetes POC over time (baseline, 9 months, 18 months)
- Virtually no patient dropout
  - Patient outcomes were obtained retrospectively after the end of the study period on a random sample of patients from each practice using chart review
  - Eligibility criteria included having a visit to the practice sometime during the study period
- Covariate selection is important
  - One approach is to include all covariates associated with the outcome with p-values less than 0.15 to 0.20 in bivariate tests, along with all covariates that are clinically meaningful (e.g. gender), associated with dropout, or differ between groups

Hierarchical and longitudinal model

- Two levels of nesting: Observations are nested within patients (baseline, 9 months, 18 months) and patients are nested within practices
- The intervention effect is actually the time*arm term, which estimates how much patient trajectories over time differ for intervention vs controls
  - Standard care: education and resources only
  - CQI: practice facilitation using continuous quality improvement approach
  - RAP: practice facilitation using a reflective adaptive process
- Basic SAS program is a random intercepts model (it is also possible to include random slopes)

```
PROC MIXED DATA = epic.patient METHOD = ML noclprint covtest;

CLASS practice id arm  time racethnicity  married education comodbidity
insurance;

MODEL  POCdiabetes = female age  racethnicity  married education comorbid
insurance arm  time  time*arm / ddfm=betwithin solution;

random intercept /sub=id(practice) type = un;

random intercept /sub=practice type = un ;

Run;
```

Colorado Epic Study: effectiveness of two approaches to practice facilitation on diabetes process of care



Overall test for differences in trend: p<.001

Moderation and heterogeneity of treatment effects

- Heterogeneity of treatment effects: response to intervention varies by patient or practice characteristics
  - A differential treatment effect involves a baseline moderator variable (sometimes called effect modification)
- In the diabetes study example, contextual effects of practice culture were examined*
- Practice Culture: measured by clinician/staff survey at baseline
  - Change culture (CC): high scores are better
  - Work culture (WC): high scores are better
- We hypothesized that change culture and work culture would *moderate* intervention effectiveness on diabetes POC,
- That is, practices with higher change culture and work culture scores at baseline would respond better (i.e. improve more) to the practice facilitation intervention
- Only two study arms are shown in this example

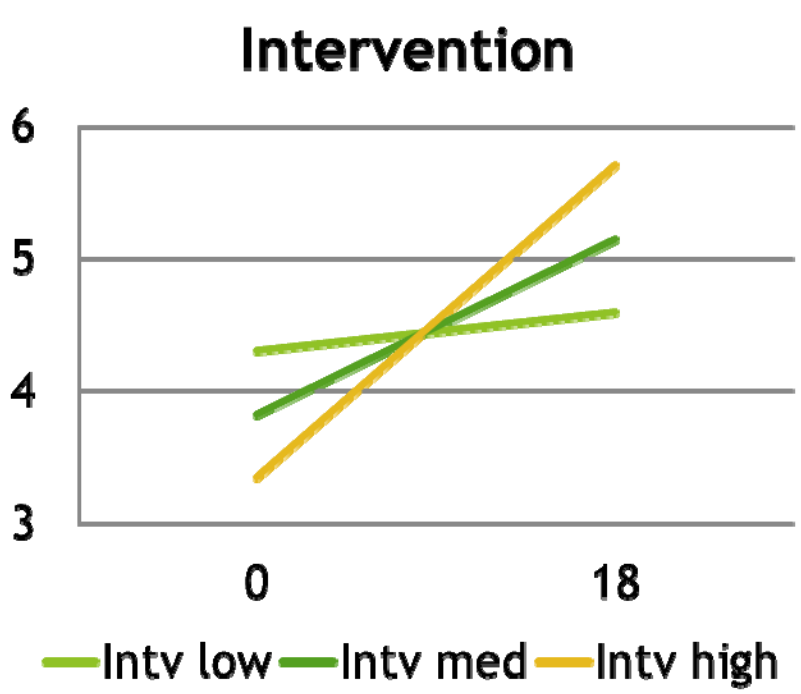
See **Dickinson LM**, et al, Practice context affects efforts to improve diabetes care for primary care patients: A pragmatic cluster randomized trial. Journal of General Internal Medicine. 2015; 30:476-82.

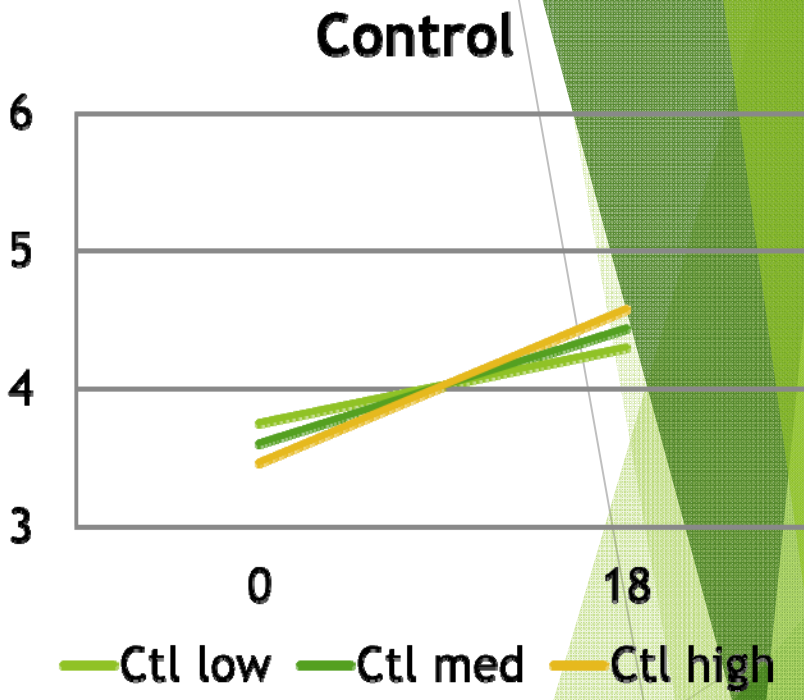
Multilevel model with practice and patient random intercepts

- Basic SAS Code (covariates not shown):
- Moderator analysis for a longitudinal study requires:
  - All main effects of interest (WC, arm, time)
  - All relevant two-way interactions
  - The three-way interaction of interest: differential intervention effect
- SAS code is the same for practice or patient moderator but underlying statistical model is different
- WC is the average practice level score on the Work Culture subscale of the Practice Culture Assessment survey

```
PROC MIXED DATA = epic.patient METHOD = ML noclprint covtest;
CLASS practice id arm   ;
MODEL  POCdiabetes =  arm  time  WC time*arm time*WC arm*WC  arm*WC*time
/ ddfm=betwithin solution;
random intercept /sub=id(practice) type = un;
random intercept /sub=practice type = un ;
Run;
```

Differential intervention effects by practice level baseline work culture scores





- Outcome is diabetes process of care
- Intervention effects differed by work culture: $p<.0001$
- Greater improvement in intervention practices with higher WC scores

Stepped Wedge Designs

- Type of crossover design that's useful when interventions likely to be effective can't be withheld from some practices
  - In practice-based research clusters are generally practices; we use the terms interchangeably here
  - Practices cross over from one condition to another at different times (0=control, 1=intervention)
  - Clusters are randomized to an intervention initiation *order*

Parallel

		Time
		1
Cluster	1	1
	2	1
	3	0
	4	0

Crossover

		Time	
		1	2
Cluster	1	1	0
	2	1	0
	3	0	1
	4	0	1

Stepped Wedge

		Time				
		1	2	3	4	5
Cluster	1	0	1	1	1	1
	2	0	0	1	1	1
	3	0	0	0	1	1
	4	0	0	0	0	1

Hussey & Hughes (2007), *Contemporary Clinical Trials,* Design and analysis of stepped wedge CRT

Stepped Wedge: Randomization and Intervention Initiation

- At the beginning of the trial, all clusters are randomized to an *order* and assigned to a step based on that order
  - In the first time block all clusters are in the control phase
- All clusters (practices) ultimately receive the intervention
  - Randomized intervention initiation order determines *when* (not if) a cluster receives the intervention By the last time block all clusters are in the intervention phase
  - By the last time block all clusters are in the intervention phase
- Traditionally, all *clusters* are recruited and enrolled at baseline and followed for the entire duration of the study (unless retrospective data are available)
- Outcomes measured for every cell (e.g. every time interval for every cluster)

Two Key Design Variations

- Repeated cross-sectional: Clusters cross over but individuals are designated as either control or intervention, depending on point of entry
  - Individuals enrolled during the control phase for that cluster are control subjects
  - Individuals enrolled during the intervention phase for that cluster are intervention subjects
  - Control and intervention groups consist of *different* individuals
- Cohort: Clusters cross over and *individuals change* from control to intervention condition at the time of the cross-over for the cluster
  - The *same individuals* are in the control and intervention phases
  - Individuals, as well as clusters, are followed throughout the entire study period
- Both can occur in the same study

Stepped Wedge Trials: sample size and analytic considerations

- Power analysis for stepped wedge is complex: involve your biostatistician early in planning phase
- Greater power than a parallel group CRT but less than traditional RCT
- Can adjust for temporal trend
- General or generalized linear mixed models can be used for analysis

- References: Hussey & Hughes (2007), *Contemporary Clinical Trials*, Design and analysis of stepped wedge CRT
- Brown CA, Lilford RJ. The stepped wedge trial design: a systematic review (link is external) BMC Med Res Methodol. 2006;6:54.
- AHRQ Stepped Wedge webinar: https://pbrn.ahrq.gov/events/advanced-methods-primary-care-research-stepped-wedge-design

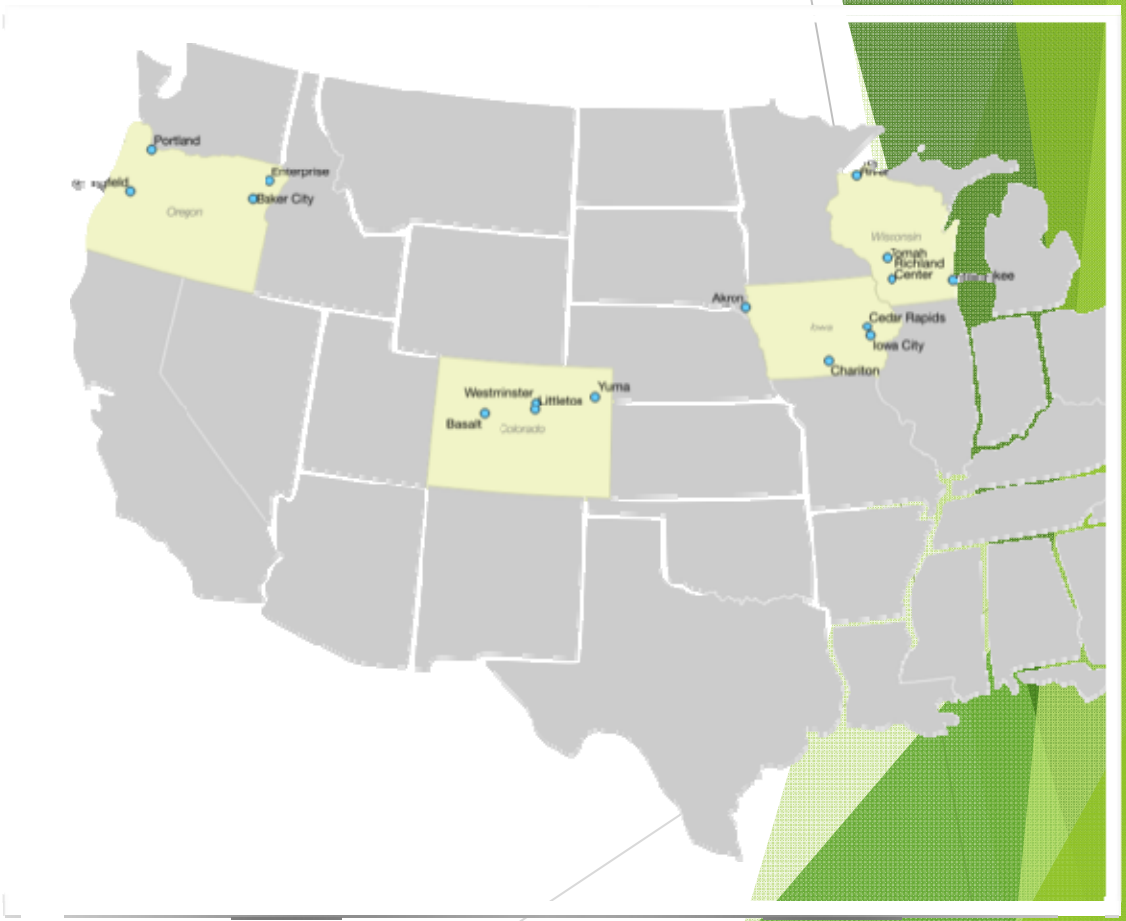
Stepped Wedge Example: Implementing Networks' Self-management Tools Through Engaging Patients and Practices (INSTTEPP)*

- Aims
  - 1. Implement the AHRQ SMS Library/Toolkit across four participating networks and 16 practices using Boot Camp Translation in a stepped-wedge design with 5 time blocks
  - 2. Assess the impact of implementation on practice staff and patients engaged in chronic care management.
  - 3. Identify the factors related to successful implementation

*funded by AHRQ

INSTTEPP

- 4 PBRN's (SNOCAP, ORPRN, WREN & IRENE)
- 16 practices
- 320 patients
- > 80 clinicians and staff



INSTTEPP Study Design

- Repeated cross-sectional for patients: Surveys (Patient Activation Measure & PACIC, self-reported health) at baseline, 1 month, and 2 months after enrollment
  - Patients enrolled during the control phase receive usual care
  - Patients enrolled during the intervention phase receive the intervention (AHRQ SMS Toolkit), with tailored delivery for each PBRN
- Cohort design for practice members: Surveys (CS-PAM & Theory of Planned Behavior) during each of the 5 time blocks
  - Clinicians/staff are in the control condition as long as the practice is in the control phase
  - Clinicians/staff cross over to the intervention phase when the practice crosses over

DM1

Slide 30

DM1 I don't understand this - there were 3 surveys for patients and 5 for clinicians

Dickinson, Miriam, 10/13/2015

Results: Patient Outcomes –
greater improvement in PACIC and self-reported health

	Survey	Control	Intervention	Differential Intervention Effect
Patient Activation Measure	1	66.72	66.07	F(,840)=0.87, p=.3515
	2	66.79	66.72	
	3	66.86	67.36	
Process of Care (from PACIC)	1	31.32	30.19	F(1,791)=16.75, p<.0001
	2	30.76	31.25	
	3	30.20	32.32	
Self-reported health (lower score is better)	1	3.17	3.35	F(1,832)=4.89, p=.0273
	2	3.16	3.25	
	3	3.16	3.16	

Adjusted for age, gender, number of chronic conditions, diabetes, chronic pain

Summary

- Key considerations in choosing a study design
  - Research question
  - Observational or experimental
  - Time frame
  - Budget and resources
  - Logistical and/or ethical concerns
  - Clustered designs
    - Parallel vs stepped wedge
    - Implications for sample size and power
    - Implications for randomization
    - Implications for measurement (especially stepped wedge)
    - Implications for data analysis

References

- Dickinson LM, Basu A. Multilevel modeling and practice-based research. Ann Fam Med. 2005 May-Jun;3 Suppl 1:S52-60.
- Murray, D. *Design and Analysis of Group-Randomized Trials*. 1998, New York: Oxford University
- Hussey & Hughes (2007), *Contemporary Clinical Trials,* Design and analysis of stepped wedge CRT
- Donner A, Klar N. *Design and Analysis of Cluster Randomization Trials in Health Research*. London, England: Oxford University Press; 2000.
- Singer J. Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *J Educ Behav Statistics*. 1998;24:323-355.
- Bryk AS, Raudenbush SW: Hierarchical Linear Models: Applications and Data Analysis Methods. Second edition ed. Newbury Park, Sage Publications, 2000.
- Dickinson, et al, Pragmatic cluster randomized trials using covariate constrained randomization: A method for practice-based research networks (PBRNs). JABFM. 2015 Sep-Oct;28(5):663-72.
- Dickinson WP, et al, Practice Facilitation to Improve Diabetes Care in Primary Care: A Report from the EPIC Randomized Clinical Trial Annals of Family Medicine. 2014; 12(1)8-16.

References

- Dickinson LM, et al, Practice context affects efforts to improve diabetes care for primary care patients: A pragmatic cluster randomized trial. Journal of General Internal Medicine. 2015; 30:476-82.
- Brown CA, Lilford RJ. The stepped wedge trial design: a systematic review. (link is external) BMC Med Res Methodol. 2006;6:54.
- AHRQ Stepped Wedge webinar: https://pbrn.ahrq.gov/events/advanced-methods-primary-care-research-stepped-wedge-design
- Fairclough DL: Design and analysis of quality of life studies in clinical trials. New York, Chapman & Hall/CRC, 2010
- Bryk AS, Raudenbush SW: Hierarchical Linear Models: Applications and Data Analysis Methods. Second edition ed. Newbury Park, Sage Publications, 2000.
- Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion intervention: the RE-AIM framework. Am J Public Health 1999, 89:1922-1927
- **www.consort-statement.org**
- Littell RC, Milliken GA, Stroup WW, Wolfinger RD: SAS system for mixed models. Cary, NC, SAS Institute, Inc., 1996