Sample Size Calculations for Pragmatic Trials

Nicolae Done

June 25, 2017

CAPER

Center for Access Policy, Evaluation and Research

QUERI
VA Quality Enhancement Research Initiative

Veterans Health Administration
Research & Development
Improving Veterans' Lives www.research.va.gov

VA HEALTH CARE Defining EXCELLENCE in the 21st Century

Objectives

1. **Analytical methods** for calculating sample size/power in pragmatic trials
   - Parallel Cluster-Randomized Trials (**CRT**)
   - Stepped-Wedge Cluster-Randomized Trials (**SW-CRT**)

2. **Factors that affect sample size/power** in pragmatic trials

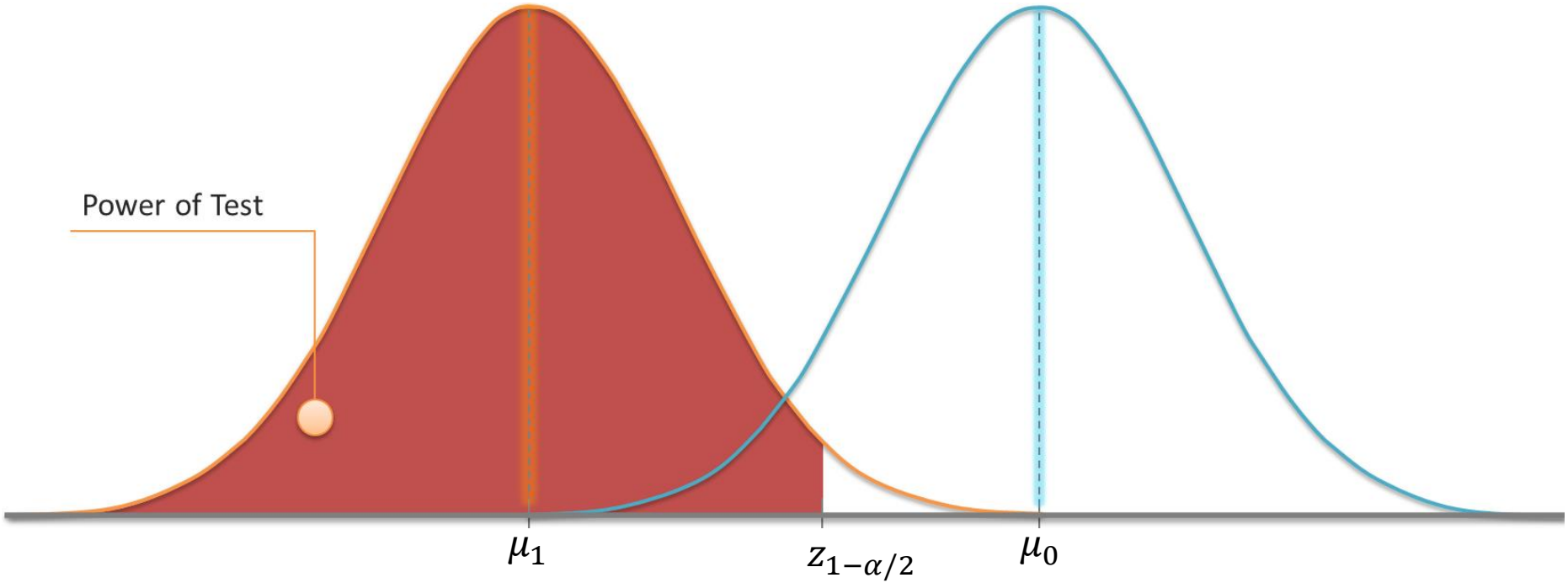
3. **Stata Packages** for calculating sample size/power

Calculating Power

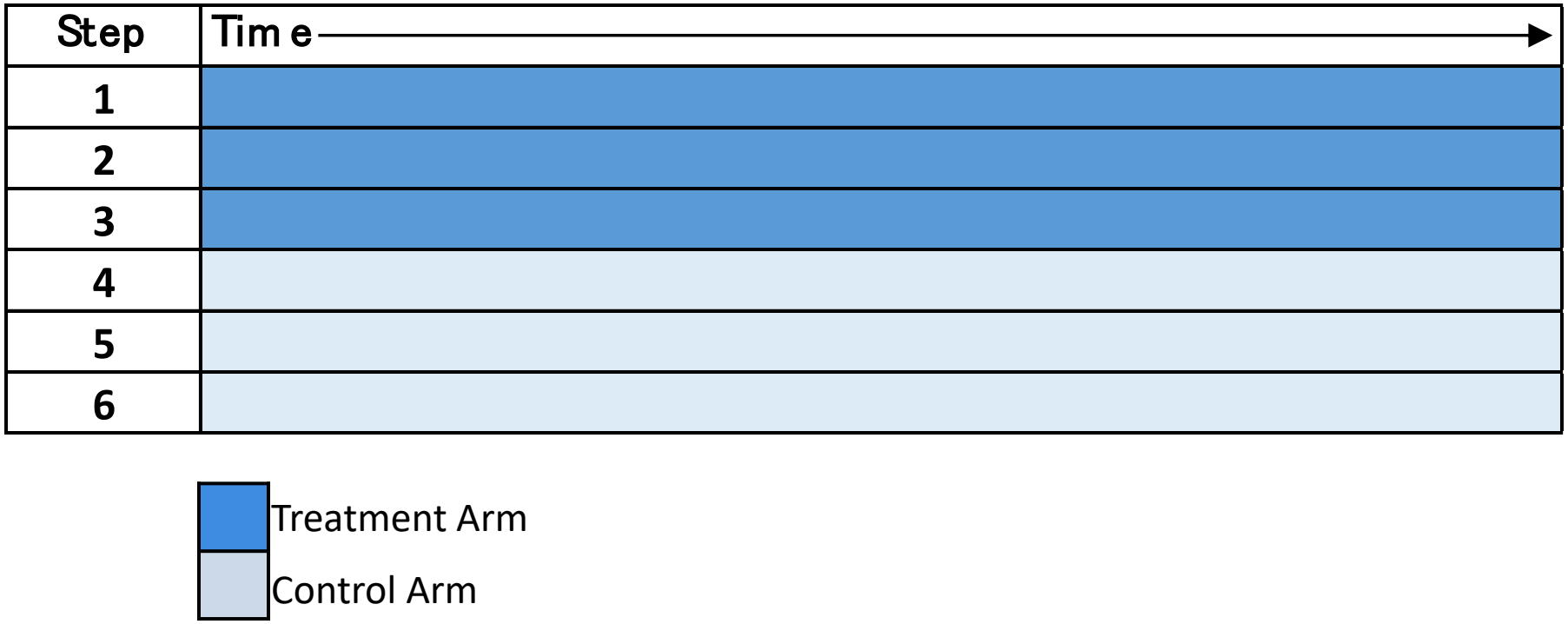
- **Analytically (asymptotic methods)**
  - Usually good enough for most practical situations
  - Relies on assumptions that may not always hold
- **Via Simulation**
  - Can accommodate almost any design
  - More complicated to perform
  - Need to specify exact data generating process

Power for a Two-Sided Test

$$1-\beta = 1 - P\left(Z < \frac{\mu_0 - \mu_1}{\frac{\sigma}{\sqrt{n}}} + z_{1-\alpha/2}\right) \quad \Leftrightarrow \quad n = \frac{\left(z_{1-\alpha/2} + z_{1-\beta}\right)^2 \times 2\sigma^2}{(\mu_0 - \mu_1)^2}$$



Parallel Cluster-Randomized Trial (CRT)



Example

Panel Management to Improve Smoking and Hypertension Outcomes by VA Primary Care Teams: A Cluster-Randomized Controlled Trial

Mark D. Schwartz, MD[1,2], Ashley Jensen, MPA[1,3], Binhuan Wang, PhD[2], Katelyn Bennett, MPH[1,3], Anne Dembitzer, MD[1,3], Shiela Strauss, PhD[4], Antoinette Schoenthaler, EdD[2], Colleen Gillespie, PhD[3], and Scott Sherman, MD[1,2]

- 20 primary care teams were randomized to **control** […] or **panel management support plus education**

- Trial duration: 8 months

Most Common CRT Assumptions

- Completely randomized design (no stratification/matching)
- Continuous or dichotomous outcomes
- Two study arms
- Equal allocation

The Design Effect (DE) for Cluster Randomized Trials

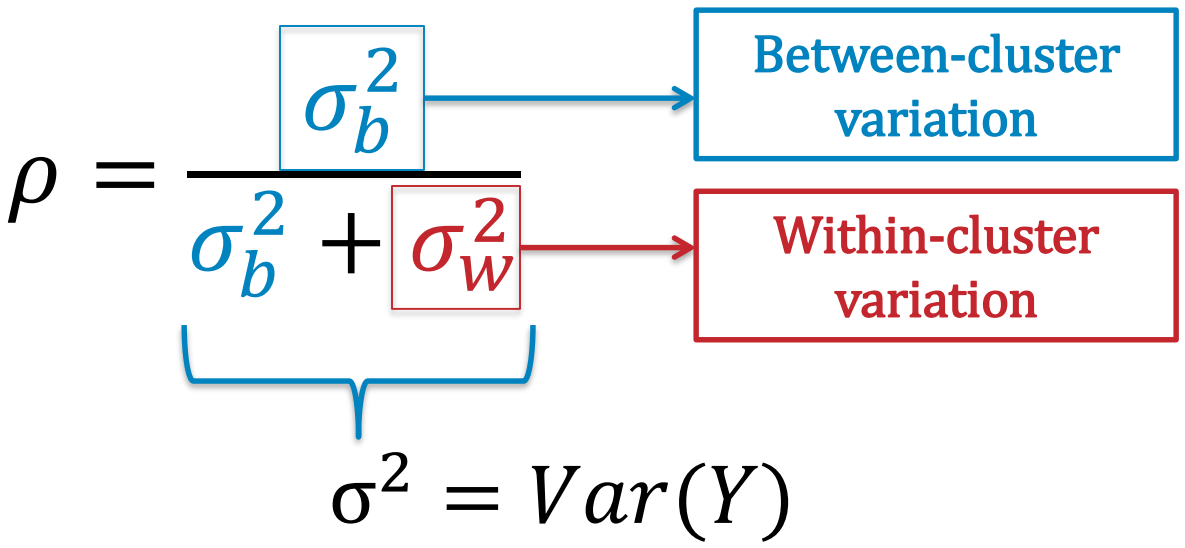
- A factor that quantifies the loss in information due to cluster randomization:

Individually Randomized Trial (*n* individuals)	$n = \frac{\left(z_{1-\alpha/2} + z_{1-\beta}\right)^2 \times 2\sigma^2}{(\mu_0 - \mu_1)^2}$
Cluster Randomized Trial (*k* clusters of size *m*, so *N=km*)	$N = km = n \times \underbrace{[1 + (m-1)\rho]}_{\textbf{DE}}$

Intracluster Correlation Coefficient (ICC)

$$\rho = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_w^2}$$

Intracluster Correlation Coefficient (ICC)



Estimating ICC Using Baseline Data

- Standard One-Way ANOVA
- Post-estimation following a random intercept regression model
- Other methods

Wu, S., Crespi, C.M., Wong, W.K., 2012. Comparison of Methods for Estimating the Intraclass Correlation Coefficient for Binary Responses in Cancer Prevention Cluster Randomized Trials. Contemp Clin Trials 33, 869–880. doi:10.1016/j.cct.2012.05.004

Parameters Required *A Priori*

- Desired statistical power $(1 - \beta)$ and significance level (α)
- Minimally significant effect size (Δ)
- Standard deviation (σ) for <u>continuous</u> outcome **OR**
- Control arm proportion for <u>dichotomous</u> outcome

- Estimated ICC (ρ)
- Cluster size (m)

How about Unequal Cluster Sizes?

- **A ~12% inflation in sample size is usually conservative**
- Express variation in cluster sizes using $\overline{m}$ and CV_m (coefficient of variation)
- Adjust for loss in efficiency using $k_{adj} = \frac{k}{1-CV_{\overline{m}}^2\ \delta(1-\delta)}$, where $\delta = \frac{\overline{m}\rho}{\overline{m}\rho+(1-\rho)}$

$$CV_m < 0.7 \text{ and } \max[\delta(1-\delta)] = 0.25 \Rightarrow \boldsymbol{k_{adj} \geq 0.88 \times k}$$

Van Breukelen & Candel (2012) Comments on 'Efficiency loss because of varying cluster size in cluster randomized trials is smaller than literature suggests'. *Statistics in Medicine*, 31(4): 397-400

Power for CRT Designs in Stata

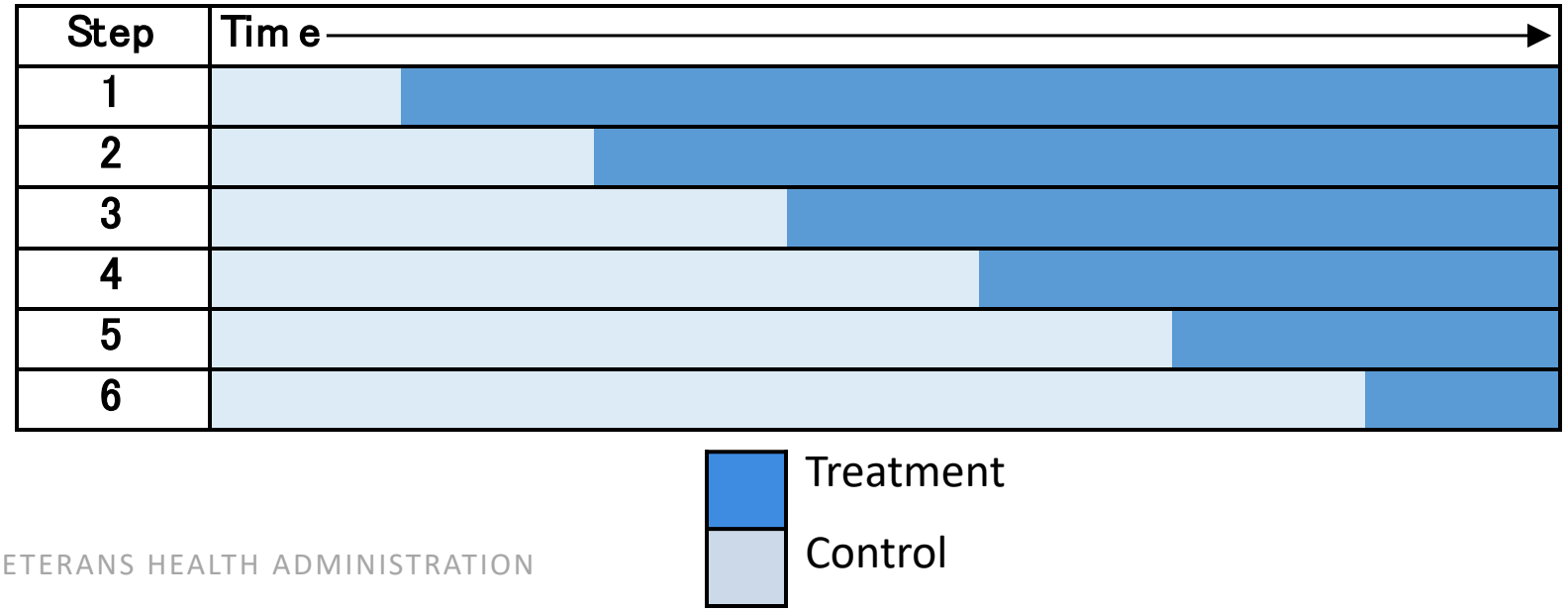
```
clustersampsi [, options]
```

options	Description
Main	
power	specify a power calculation
detectabledifference	specify a detectable difference calculation
samplesize	specify a sample-size calculation (default)
binomial	specify a comparison of proportions (comparison of means is the default)
rates	specify a comparison of rates (comparison of means is the default)
rho(#)	intraclass correlation coefficient
cluster_cv(#)	coefficient of variation of outcomes; default is **cluster_cv(0)**

Available via:
ssc install clustersampsi

Stepped-Wedge CRT

- Balanced, Complete, Cross-Sectional:
  - All *k* clusters start in control condition, then cross to intervention sequentially
  - Outcomes measured in each of *T=t+1* periods
  - Cross-sectional samples of *m* individuals/cluster in each period



Example

- VA Telederm Trial
- 38 VA facilities randomized to 6 steps
- Intervention = a new mobile app for teledermatology + training and facilitation
- Outcomes = consult completion times, % telederm adoption
- Measurement done passively via extraction from VA CDW

Design Effect for Stepped-Wedge Cluster Randomized Trials

- Total required sample size:

$$km(t+1) = n_{ind} \times (t+1) \times DE$$

$$DE = \frac{1+\rho(tm+m-1)}{1+\rho(\frac{tm}{2}+m-1)} \times \frac{3(1-\rho)}{2(t-\frac{1}{t})}$$

- *t* needs to be specified in advance by considering logistics, planned study duration, and available *m*

Hussey, M.A., Hughes, J.P., 2007. Design and analysis of stepped wedge cluster randomized trials. Contemporary Clinical Trials 28, 182–191. doi:10.1016/j.cct.2006.05.007

Power for SW-CRT Designs in Stata

```
steppedwedge [, options]
```

options	Description
Main	
power	specify a power calculation
detectabledifference	specify a detectable-difference calculation
binomial	specify a comparison of proportions (comparison of means is the default)
rates	specify a comparison of rates (comparison of means is the default)
alpha(#)	the significance level of the test; default is **alpha(0.05)**
beta(#)	the power of the test; default is **beta(0.8)**

Available via:
ssc install steppedwedge

Acknowledgments and Author Affiliations

- **Nicolae Done, Boston University School of Medicine and CAPER**
- Julia C. Prentice, Boston University School of Medicine, School of Public Health, and CAPER
- Melissa Garrido, Icahn School of Medicine at Mount Sinai, Bronx VA GRECC and PEPReC

- Funded by grant number **PEC 15-467** from QUERI
- All opinions expressed are those of the authors and do not necessarily reflect the official position of the U.S. Department of Veterans Affairs, Boston University, or Northeastern University.

- **Correspondence**: nicolae.done@va.gov

Appendix – Minimal Example Code

clustersampsi, binomial samplesize p1(0.3) p2(0.33) m(40) rho(0.05) size_cv(0.9)

steppedwedge, binomial power complete(1) p1(0.3) p2(0.33) m(40) k(3) rho(0.05) alpha(0.05) steps(10) vartotal(1) dm(1)