

# **Advantages and Disadvantages of Various Randomized Clinical Trial Designs**

Brian Leroux

[leroux@uw.edu](mailto:leroux@uw.edu)

University of Washington

**IADR/AADR Satellite Symposium:  
Updates in Clinical Research Methodology  
March 18, 2013**

Supported by NIDCR grants DE016750, DE016752.

# Outline

- Introduction
  - Why randomize?
  - History and use of RCTs
  - Phases of RCTs
- Clinical trial designs
  - Completely randomized design
  - Stratified design
  - Cross-over design, split-mouth design
  - Cluster randomized design
  - Cluster-crossover design
  - Step-wedge design
- How to choose a design - Examples

# Why Randomize?

Random assignment of patients to treatments provides the strongest possible basis for inference about treatment effects.

To take advantage of this we perform an **Intent-to-Treat Analysis:**

patients are analyzed according to their random treatment assignment, i.e., the *intended* treatment, not the treatment actually received.

# History of RCTs

- 1880s – First randomized experiments in psychic research\*
- 1925 – Fisher's lady tasting tea^
- 1948 – First published medical RCT&
- 1968 – One of the earliest dental RCTs#
- # RCT publications:
  - in Medline: 340,000
  - in Dental journals: 10,000

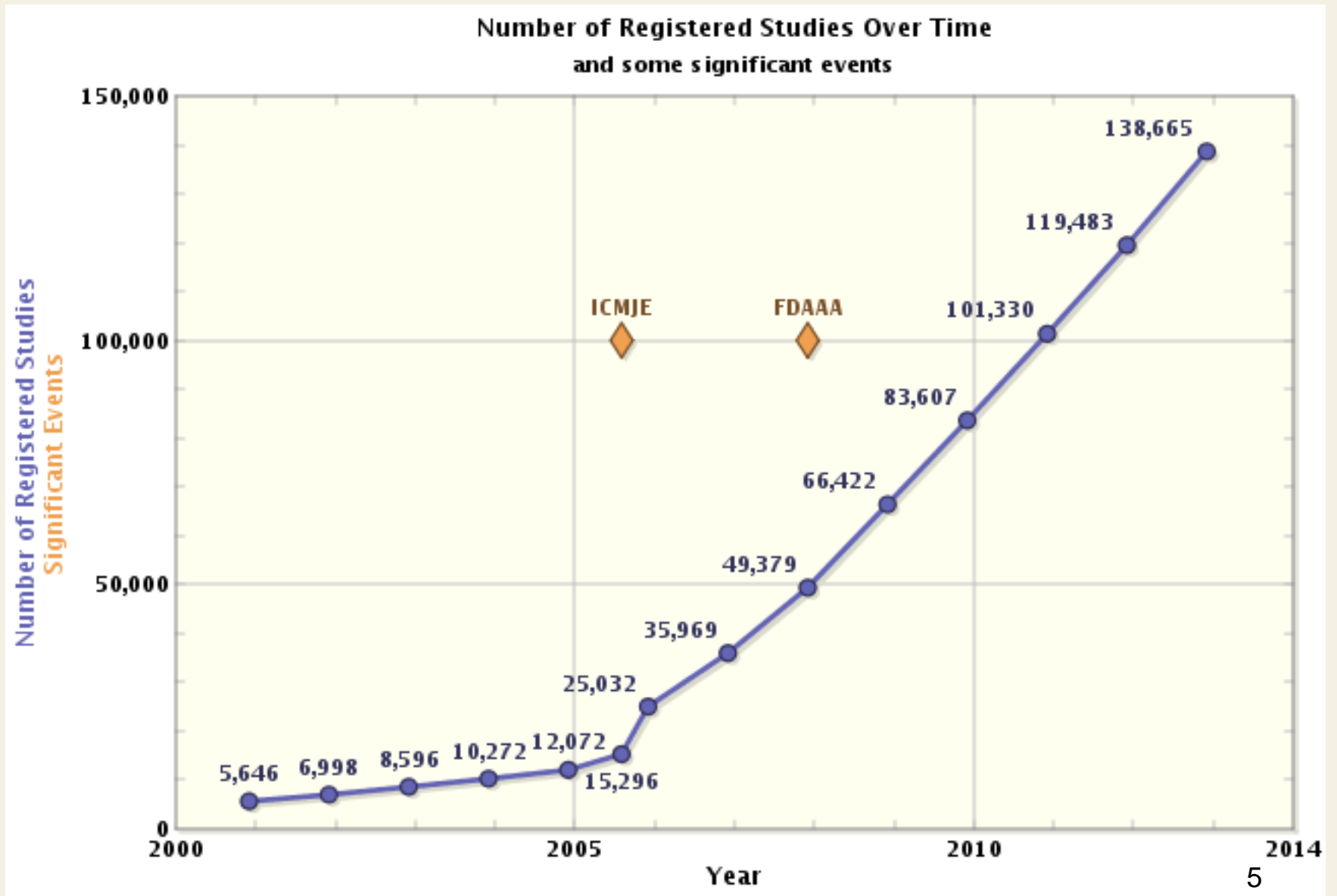
\*Hacking, I (1988). *ISIS* 79(3):427-451.

^Fisher, R.A. (1925). *Statistical Methods for Research Workers*.

& Medical Research Council (1948). *BMJ* 2:769-782.

# Ramfjord et al. (1968). *J Periodontol.* 39(3):167-

# Studies Registered in ClinicalTrials.gov



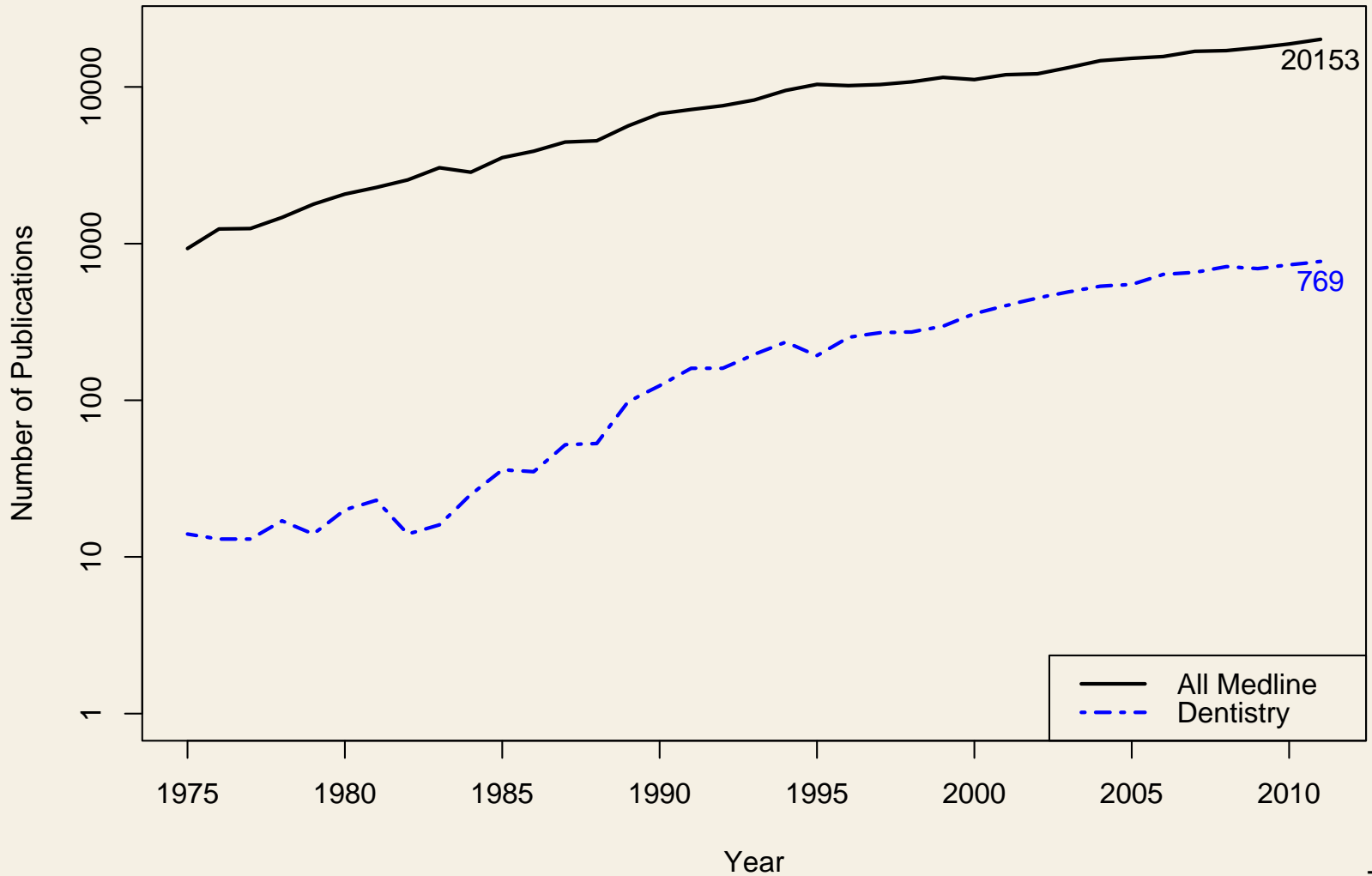
# # Studies in ClinicalTrials.gov by Topic

Topic	# Studies*
Gingivitis	377
Periodontitis	321
Oral Cancer	248
Caries	206
Dental Implants	131
Restorative Dentistry	98
TMD	90
Maxillofacial Surgery	84
Orthodontics	75
Cleft Lip/Palate	48
Endodontics	36

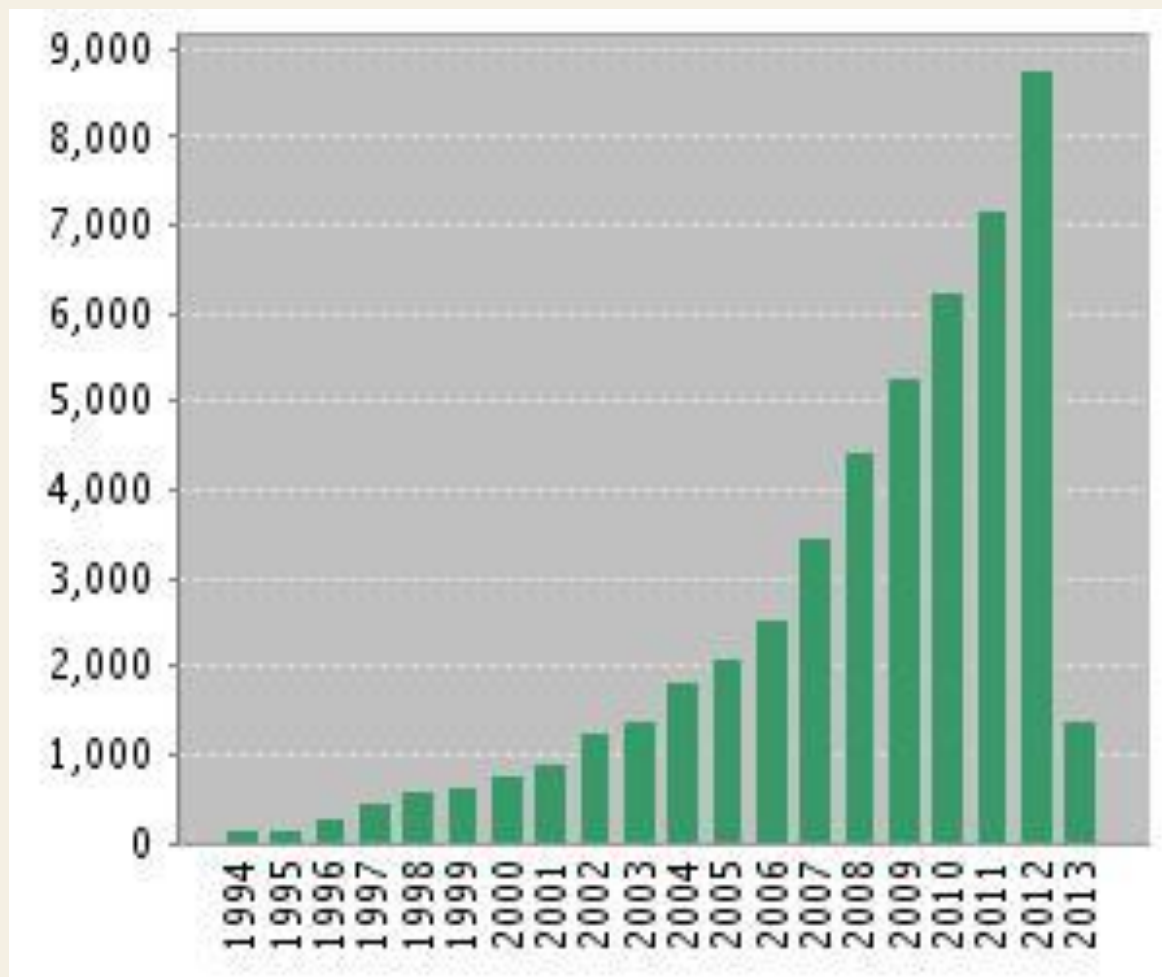
\* Not necessarily unique studies.

# Growth of RCTs

# RCT Publications 1975-2011



# Citations of Dental RCTs\*



\* Web of Science, articles published 1975-2011. Average citations per article: 15, h-index: 74.



# Average # Citations by Field

Field	# RCT Articles	Average # Citations per Article
Ophthalmology	4445	22
Geriatrics	4365	20
Transplantation	3656	22
<i>Dentistry</i>	<i>3317</i>	<i>15</i>
Nursing	3287	10
Microbiology	3039	32
Health Policy	3010	17
Dermatology	2994	18
Psychology	2739	24
Reproductive Biology	2603	19

# Phases of Clinical/Translational Research

- Phase “0” – basic science & animal studies
- Phase 1 – small trials to explore biological activity and safety (possibly unrandomized)
- Phase 2 – medium-sized trials to explore treatment effect on surrogate outcomes
- Phase 3 – large trials to provide definitive evidence of treatment effect on clinical outcomes
- Phase “3.5” – meta-analyses of RCTs
- Phase 4 – observational studies in real world

# RCT Designs

The defining feature of an RCT design is the way in which the random assignment of treatments is performed.

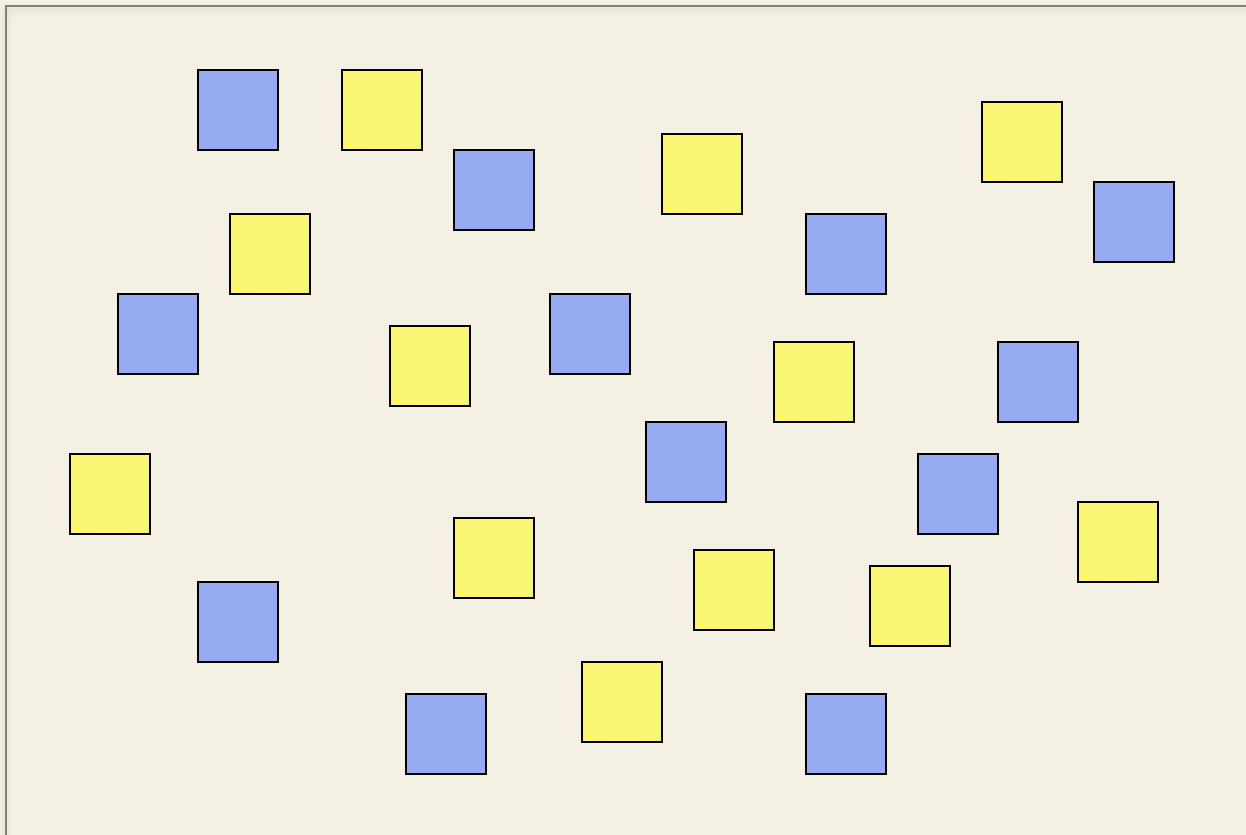
There are many ways to randomize!  
by patient, by clinic, by tooth, ....

# Main Types of RCT Designs

- Completely Randomized
- Stratified
- Cross-over
- Cluster-randomized
- Cluster-crossover
- Step-wedge

# Completely Randomized RCT

Recruit sample of patients and randomize



# Pros and Cons of Completely Randomized Design

- + Relatively simple to execute
- May be less efficient than other designs

## Example – Phase I RCT of Glycine Powder Air Polishing in Periodontal Pockets\*

- Comparison of sub-gingival air polishing with glycine powder vs scaling and root planing
- Completely randomized (n=15 per group)
- By chance, there was a surprising imbalance between groups in gender, which didn't appear to affect results (but is a good example of why stratification is useful)

\* Flemmig et al. (2012). J. Periodontology 83(4):444-452.

# Sample Size Calculation for Completely Randomized Design

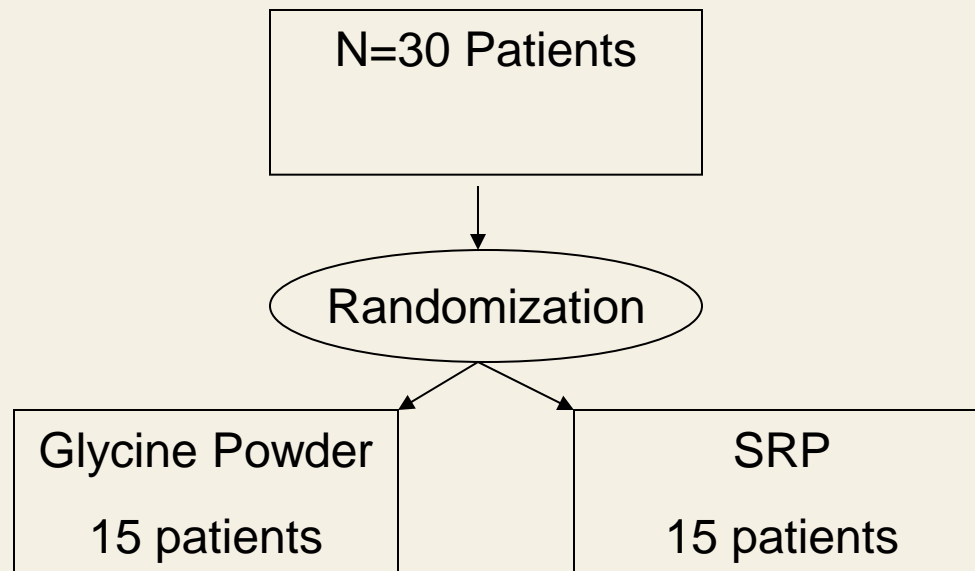
- Outcome = change in average total bacterial count on log scale, in moderate-deep pockets
- Based on previous similar study
  - hypothesized difference of 1 (on log scale)
  - SD = 0.88
- $N = 2 \times [(1.96 + 0.84) \times 0.88/1]^2 = 12$  patients per group for power 80%, type I error 0.05\*
- Enroll 15 per group to allow for missing data and correct for small sample<sup>^</sup>

\*General formula:  $N = 2 \times [(z_a + z_b) \times \sigma/\Delta]^2$

<sup>^</sup> Formula for N is a large-sample approximation.

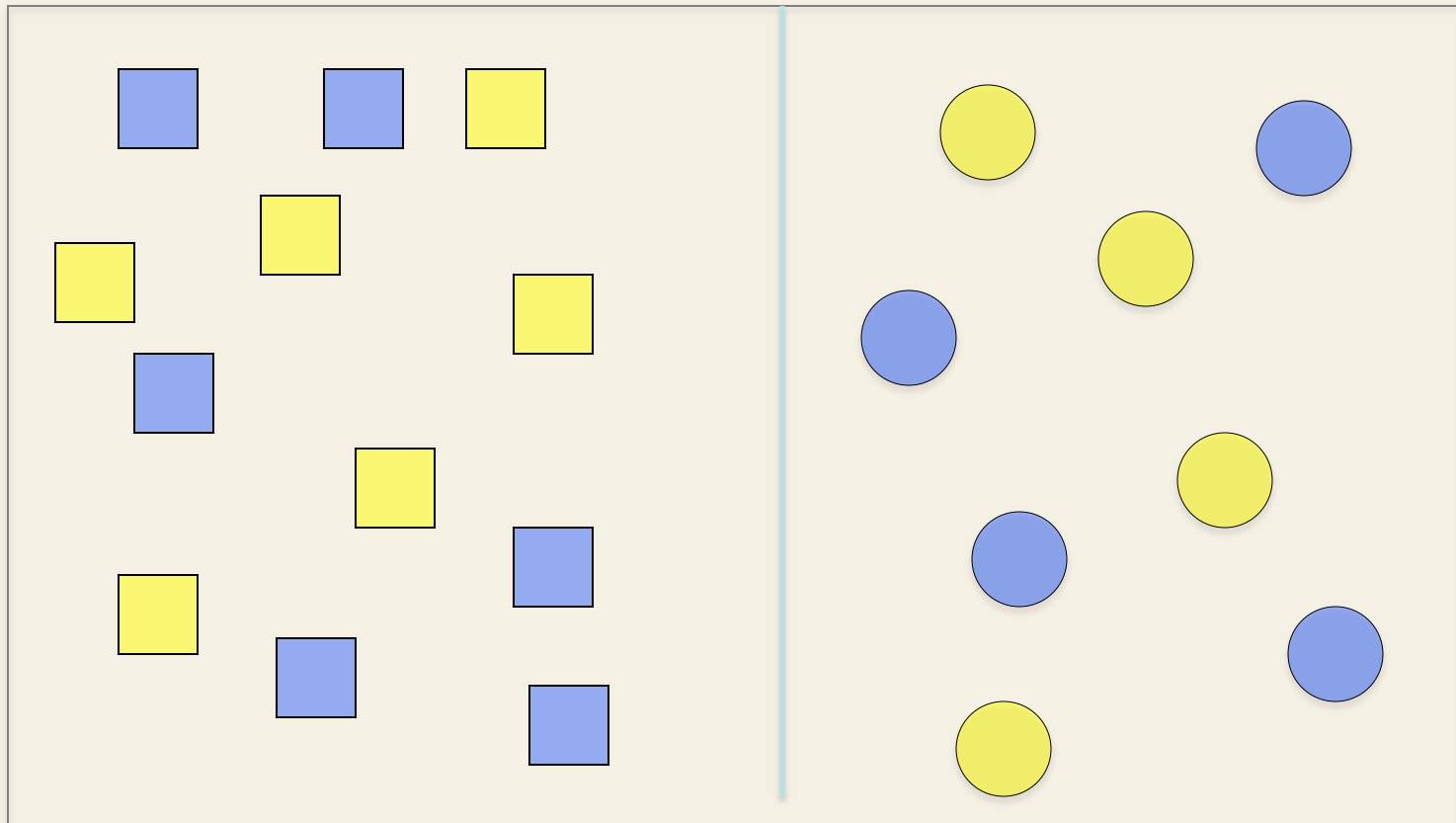


# Completely Randomized Design for Glycine Powder RCT



# Stratified RCT

Recruit samples of patients within strata and **randomize within** strata.



# Pros and Cons of Stratified Randomized Design

- + Gain in precision – compare treatments within strata, so variation between strata is eliminated
- + Treatment groups guaranteed to be fairly well balanced on stratum characteristics
- + Can compare treatment effects across strata
- Can be difficult to implement – need to know stratification variables before randomization

# Example – Treatment of White Spot Lesions for Orthodontic Patients\*

- Comparison of
  - 1) MI Paste Plus
  - 2) Fluoride varnish
  - 3) Regular home care
- Randomization stratified by dental practice and time since band removal (< 2 months versus 2 months or more prior to randomization)

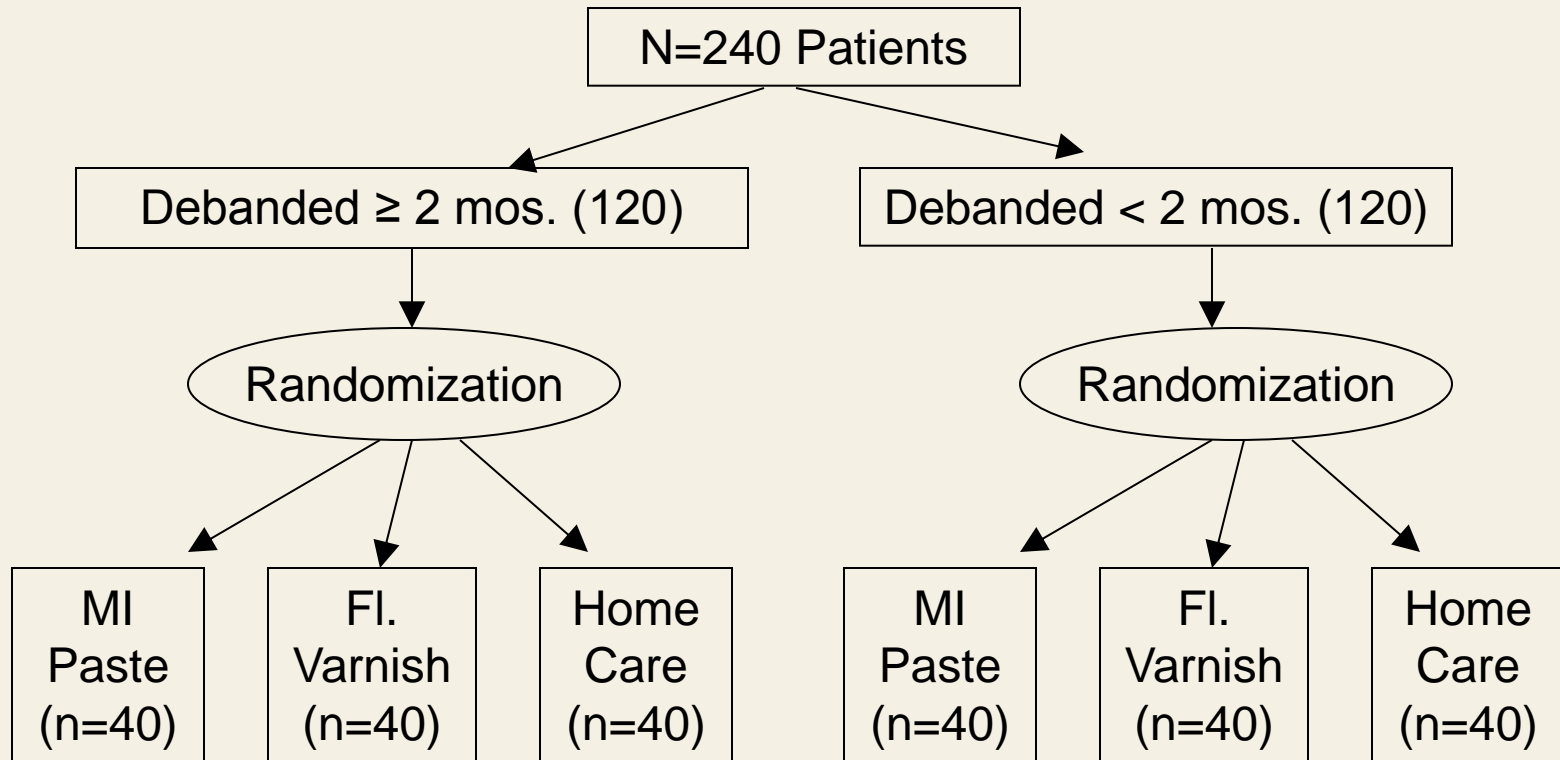
\* Huang GJ et al. (2013). *Amer J Orthodont Dent Orthoped* 143(1):31-41. (ClinicalTrials.gov # NCT 01059058)

# Sample Size Calculation for White Spot Lesions Trial

- Outcome = 8-week % improvement in WSL assessed by blinded examiners using VAS
- Difference desired to detect = 20%
- Within-stratum SD = 25%\*
- Type I error 0.025, power=80% for pairwise comparisons
- $N = 2 \times [(2.24 + 0.84) \times 25/20]^2 = 30$
- Enrollment target: 40 per group
  - Aim for 40 per group for both  $< 2$  and  $\geq 2$  months, to be able to detect treatment effects separately

\* Note: The overall SD may be higher than this.

# Stratified Randomized Design



# Results for WSL RCT

No evidence for difference between treatments.

	<b>MI Paste</b>	<b>Fl. Varnish</b>	<b>Home Care</b>
	(N=34)	(N=40)	(N=41)
Improvement*, mean (SD)	21.1 (21.6)	28.5 (25.9)	27.3 (22.7)

95% CI for difference:

MI vs Home Care: (-19.1, 6.7)

Fl. Varnish vs Home Care: (-11.8, 14.2)

\*As assessed by expert panel. Similar results for lay panel, self-assessment and objective measure.

# Crossover Design

All patients receive both treatments. Randomize patients to order of treatments.

	Period 1	Period 2
Patient 1	C	T
Patient 2	C	T
Patient 3	T	C
Patient 4	C	C
Patient 5	C	T
Patient 6	T	C
Patient 7	T	C
Patient 8	C	T



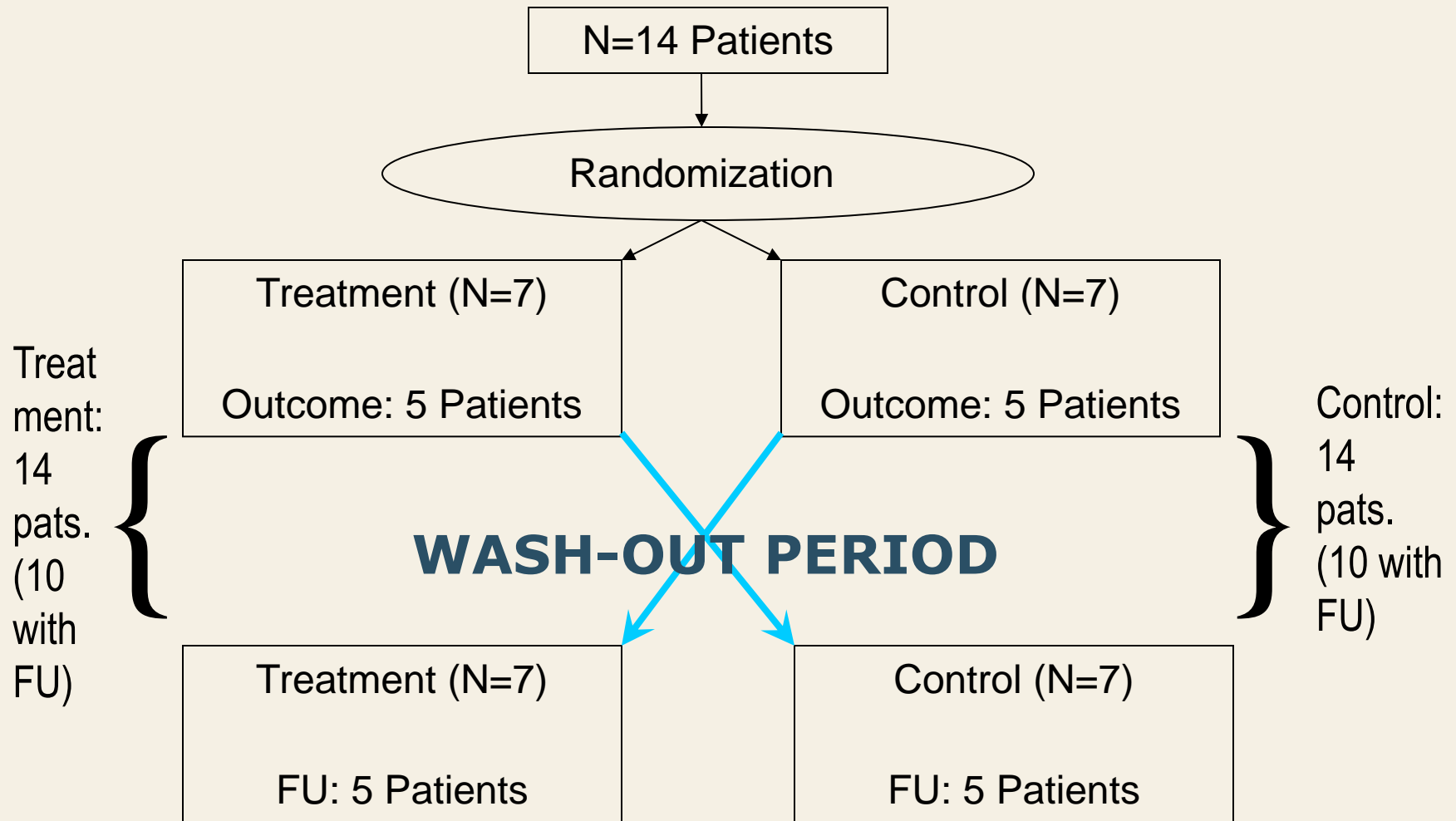
# Pros and Cons of Crossover Design

- + Gain in precision – compare treatments within patients, so variation between patients is eliminated
- Susceptible to carry-over effects (lasting effects of one treatment on outcome in the following period), which can invalidate the results

# Sample Size Calculation for Cross-over Design

- The design effect (DEFF) tells you the reduction in required sample size due to correlation between measures on same patient
- Design effect =  $1 - \rho$ 
  - $\rho$  = correlation between outcomes on same patient
- Required N = Naïve N  $\times$  (1 -  $\rho$ )
- Example (Glycine powder air polishing):  
 $N = 12 \times (1 - 0.2) = 9.6$
- So need 10 patients with 2 periods each
  - But enroll 14 patients to account for dropout

# Crossover Design



Need follow-up on at least 10 patients (2 periods each)

# Split-Mouth Design

All patients receive both treatments in different teeth. Randomize teeth within patients.

Warning: carry-over effects may invalidate results.\*

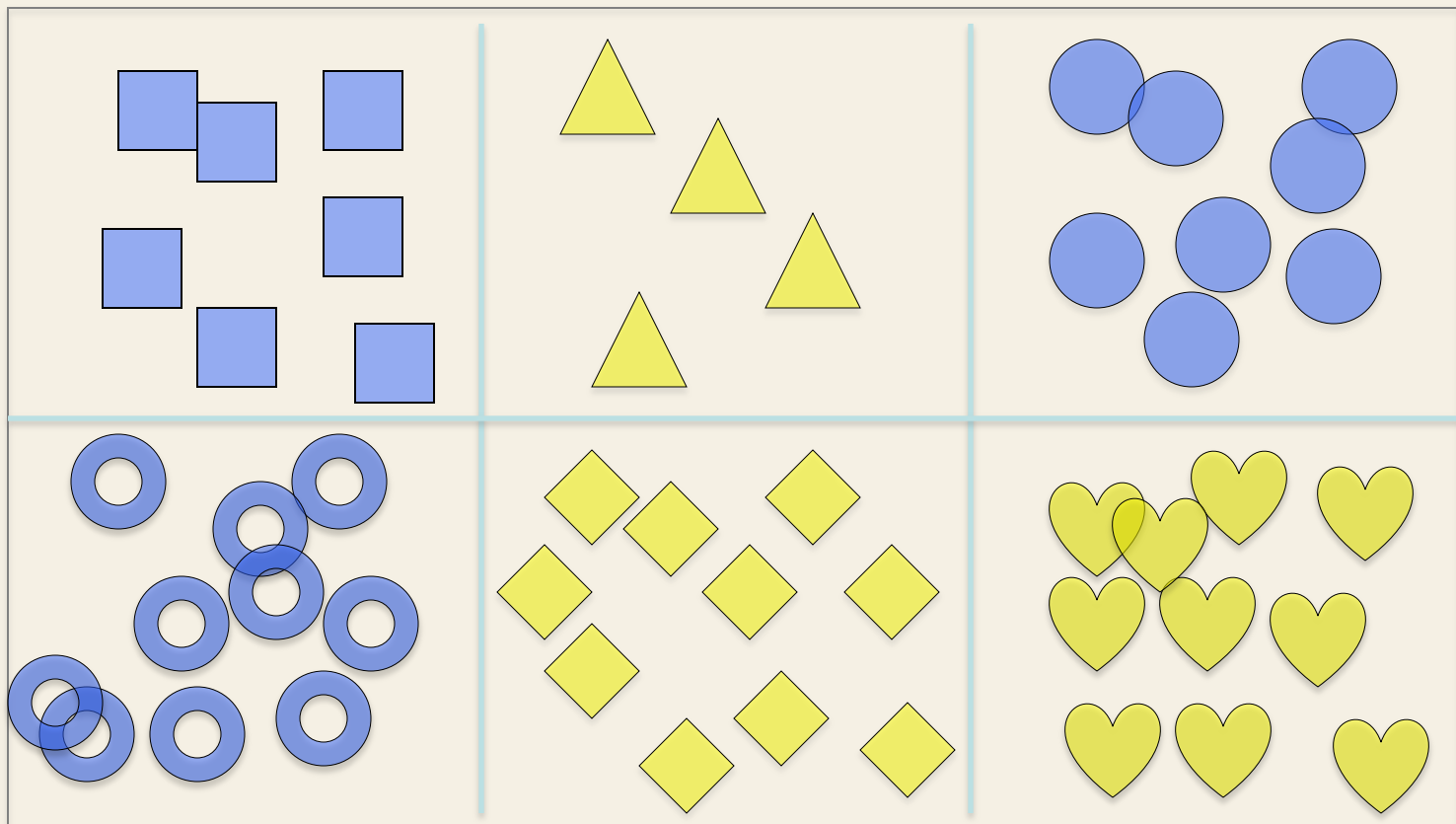
	<i>Tooth 1</i>	<i>Tooth 2</i>
Patient 1	<b>C</b>	<b>T</b>
Patient 2	<b>C</b>	<b>T</b>
Patient 3	<b>T</b>	<b>C</b>
Patient 4	<b>C</b>	<b>C</b>
Patient 5	<b>C</b>	<b>T</b>
Patient 6	<b>T</b>	<b>C</b>
Patient 7	<b>T</b>	<b>C</b>
Patient 8	<b>C</b>	<b>T</b>

\*Lesaffre et al. (2009). *Statist Med* 28(28):3470-3482.

Hujoel (1998). *Comm Dent Oral Epi* 26(2):85-86.

# Cluster-Randomized RCT

Recruit a sample of **clusters of patients** (clinics, communities, etc.) and **randomize the clusters**.



# Pros and Cons of Cluster-Randomized Design

- + Prevents contamination between individuals assigned to different treatments (patient or provider)
- + Useful for group-level interventions
- + Useful when individual randomization is impossible
- Can be much less efficient than individual-randomized designs

# Example: PRECEDENT Pulp Capping RCT\*

- MTA vs CaOH for direct pulp capping
- Reasons for cluster randomization:
  - Emergency procedure so not feasible to randomize individual patients
  - Practice effects could contaminate results if randomized individual patients

\* ClinicalTrials.gov # NCT00812877

# Sample Size Calculation for Pulp Capping RCT

- Outcome = failure within 2 years
- Failure probability 0.15 vs. 0.30
- Type I error 0.05, power = 90%
- Ignoring clustering:

$$N = [1.96\{2(.225)(1-.225)\}^{1/2} + 1.28\{.15(1-.15)+.3(1-.3)\}^{1/2} / 0.15]^2 = 161^*$$

\*In general:  $N = ( [z_a \{2p(1-p)\}^{1/2} + z_b \{p_1(1-p_1)+p_2(1-p_2)\}^{1/2} / [p_1-p_2] )^2$



# Accounting for Clustering (Design Effect)

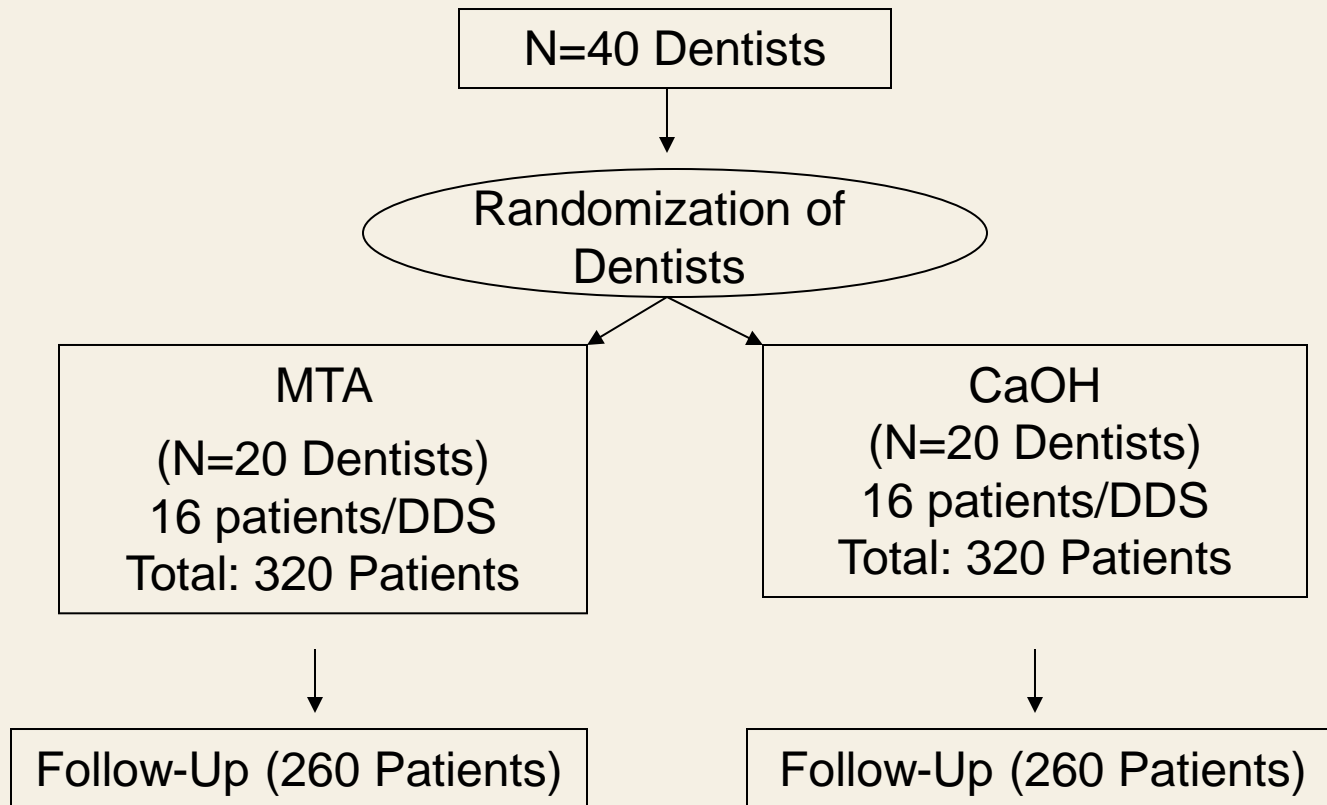
- The design effect (DEFF) tells you how much to inflate N to account for clustering
- $DEFF = 1 + (m-1) \times ICC$ 
  - $m$  = average # patients per group (assume 13)
  - ICC = intraclass correlation = 0.05 (see next slide)
  - Design effect =  $1 + (13 - 1) \times 0.05 = 1.6$
- $N = \text{Naïve } N \times DEFF = 1.6 \times 161 = 258$
- Need 20 dentists per group (13 patients/dentist) to have  $20 \times 13 = 260 \geq 258$  patients.
- So enroll 16 patients per office (allow loss of 3)

\* Murray, D. *Group Randomized Trials*.

# How to estimate the ICC

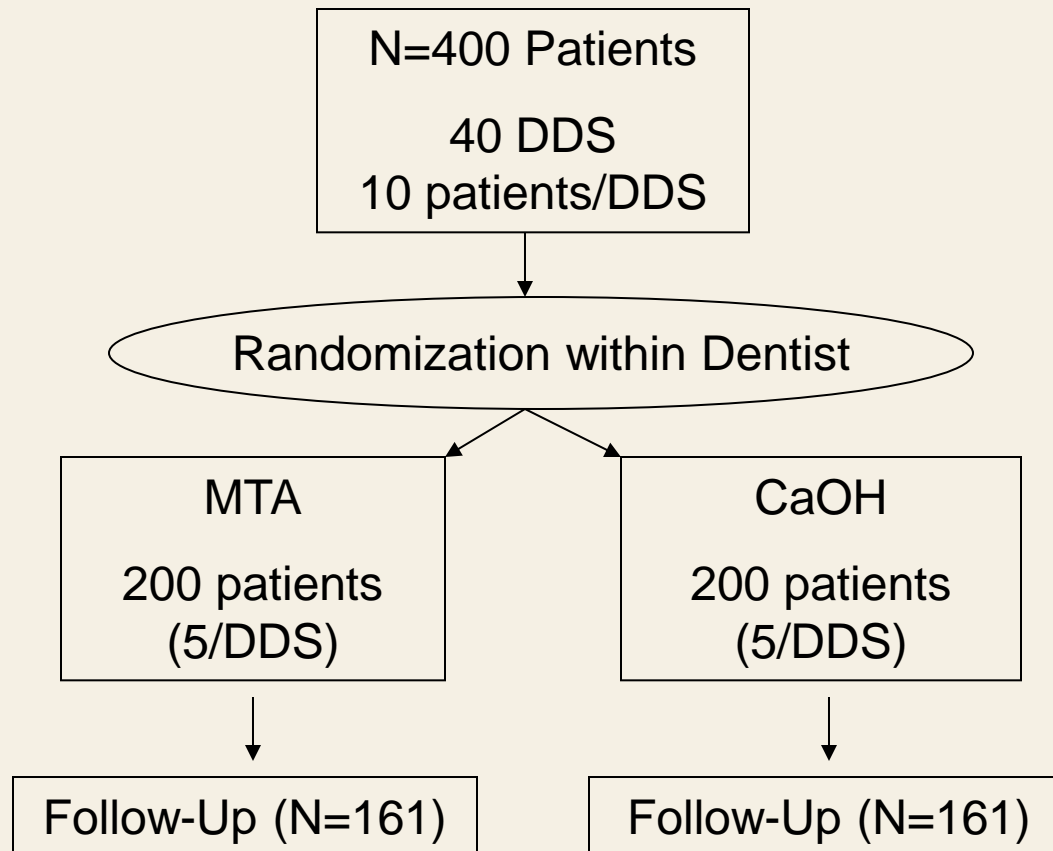
- ICC = correlation between outcomes on 2 patients in same dental practice
- Prior PRECEDENT studies showed that ICC varies greatly with type of outcome:
  - Process variables (e.g., use of amalgam): ICC sometimes quite large, e.g., 0.3
  - Disease outcome variable: ICC often in range 0.01 to 0.05
- Used ICC=0.05 for planning pulp cap trial

# Cluster-Randomized Design for Pulp Cap Trial



Need follow-up on total of 520 patients, 260 per treatment.

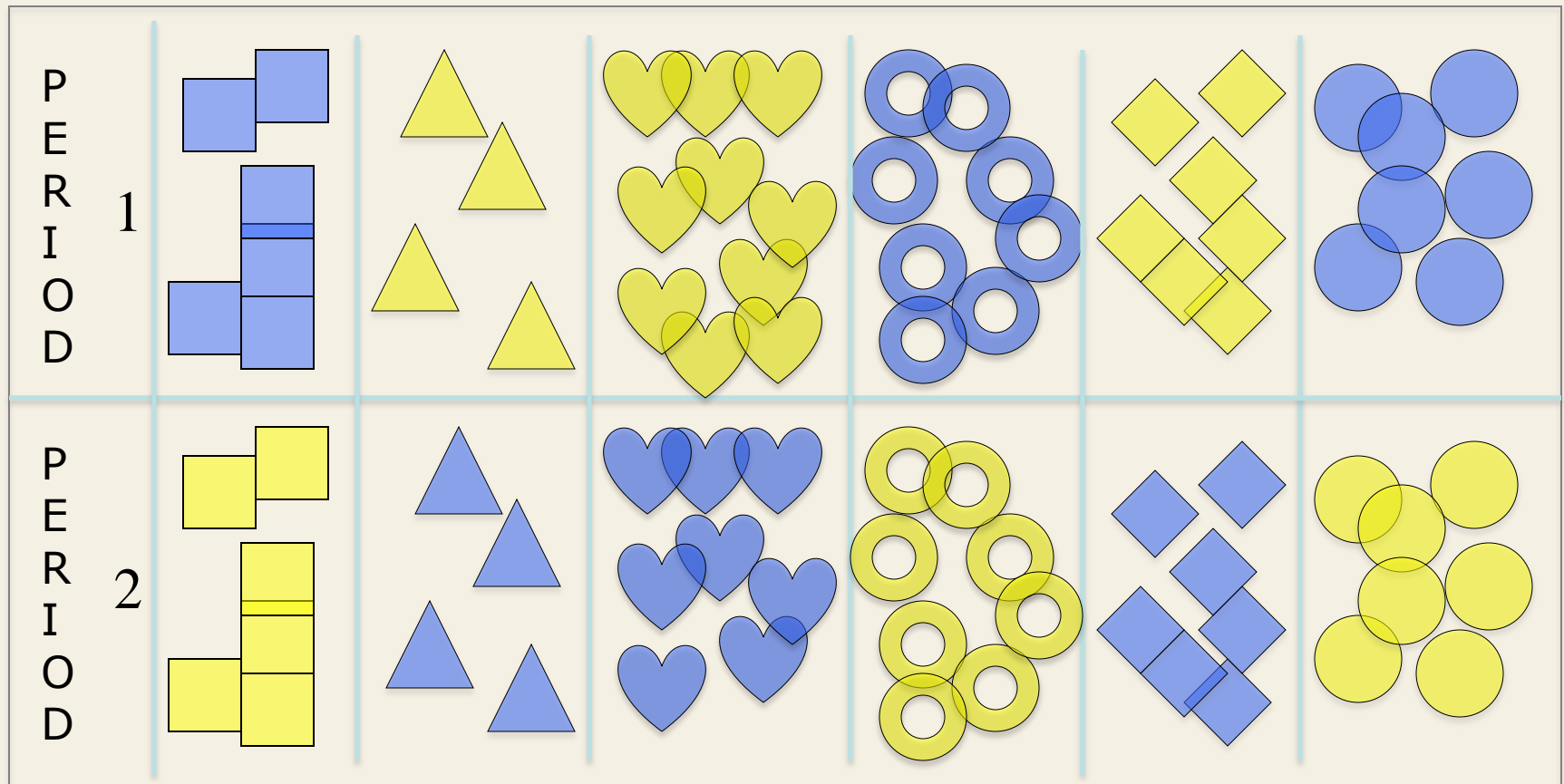
# A Stratified Design for Pulp Cap Trial (if it was feasible)



Need follow-up on total of 322 patients, 161 per treatment<sub>36</sub>

# Cluster-Crossover RCT

Clusters are randomized to treatments, outcomes measured, then **clusters cross-over** to other treatment.



# Pros and Cons of Cluster-Crossover Randomized Design

- + Recovers loss of efficiency due to clustering
- Susceptible to carry-over effects (like patient crossover design)
- Doubles length of trial

# Example: PRECEDENT Pulp Capping RCT

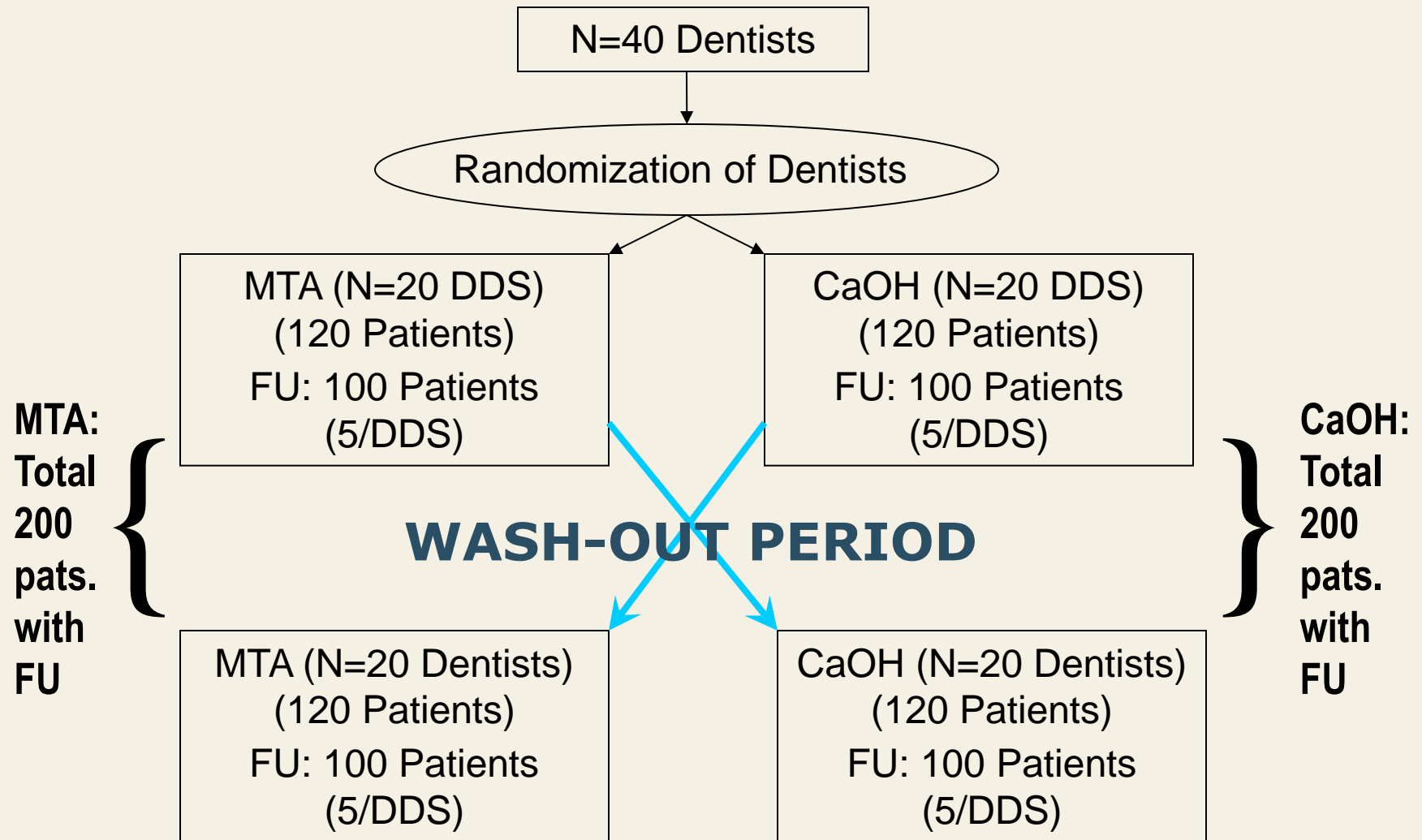
- Cluster-crossover maintains advantages of the clustered design but recovers some of the loss of power due to clustering of patients by practice
- Sample size calculation now depends on two correlations:
  - ICC = correlation between outcomes in the same practice in the same period
  - IPC = correlation between outcomes in the same practice but different periods

# How big is the IPC?

- IPC = correlation between outcomes on 2 patients in same dental practice in different treatment periods
- Reasonable to assume  $0 < IPC \leq ICC$
- Greatest benefit from cluster-crossover if  $IPC=ICC$
- Least benefit if  $IPC=0$ 
  - When planning a CC trial best to use this unless you have good prior estimate, eg, from a pilot trial



# Cluster-Crossover Design for Pulp Cap RCT



Need follow-up on 400 patients, 200 for each treatment

# Step-Wedge RCT

Randomize clusters to start-time of intervention.

	Time 1	Time 2	Time 3	Time 4	Time 5	Time 6
Cluster 1	C	C	C	T	T	T
Cluster 2	C	T	T	T	T	T
Cluster 3	C	C	C	C	T	T
Cluster 4	C	C	T	T	T	T
Cluster 5	T	T	T	T	T	T
Cluster 6	C	C	C	T	T	T
Cluster 7	C	C	C	C	C	T
Cluster 8	C	T	T	T	T	T

# Pros and Cons of Step-Wedge Design

- + Similar advantages to cluster-crossover design
- + Useful if there are carry-over effects (e.g., community-based prevention, vaccine studies)
- Requires rapid enrollment and outcome assessment to control length of each period
- Susceptible to temporal confounding

# How to Choose a Design?

- Optimal design will depend on many parameters, some of which are hard to determine before you do the study!

# Example 1: Pulp Cap RCT

Design	+/-
Completely Randomized	<ul style="list-style-type: none"><li>• Not advisable for multi-center study</li></ul>
Stratified	<ul style="list-style-type: none"><li>• Low N (322) but infeasible and subject to treatment to contamination</li></ul>
Cross-over/Split-mouth	<ul style="list-style-type: none"><li>• Not feasible to find patients</li></ul>
<b>Cluster-Randomized</b>	<ul style="list-style-type: none"><li>• <b>Best because it avoids contamination although requires large N (520)</b></li></ul>
Cluster-crossover	<ul style="list-style-type: none"><li>• Moderate N (400) but takes more time</li></ul>
Step-wedge	<ul style="list-style-type: none"><li>• Takes too much time</li></ul>

# Design Choice for Pulp Cap RCT

- Cluster-randomized design is best overall although sample size required is largest
- Would be nice to do cluster-crossover but requires much more time given that pulp cap procedures are not common and need for wash-out period
- Other designs (individual patient cross-over, split-mouth, step-wedge) are not feasible

## Example 2: Rare Adverse Outcome

- Compare 2 restorative materials/techniques
- Outcome = occurrence of adverse outcome
  - Assume incidence = 4% under control vs 2% with new material
- If no clustering, need 2280 patients for 80% power with type I error 0.05 (1140 per group\*)
- Consider 100-300 practices, each enrolling on average 8-23 patients

\*  $N = [1.96\{2(.03)(1-.03)\}^{1/2} + 1.28\{.02(1-.02)+.04(1-.04)\}^{1/2} / 0.02]^2$

# Rare Outcome (100 Practices)

# Patients required for power 0.8 to detect 4% vs 2%

# DDS	ICC	Cluster-random.	CC* (cons)	CC (opt)
100	0.01	2,930	2560	2280
100	0.02	4,120	2900	2260
100	0.03	7,020	3380	2240
100	0.04	25,090	4040	2220

\*CC = cluster-crossover



# Rare Outcome (300 Practices)

# Patients required for power 0.8 to detect 4% vs 2%.

# DDS	ICC	Cluster-random.	CC (cons)	CC (opt)
300	0.01	2460	2400	2280
300	0.02	2640	2460	2280
300	0.03	2880	2520	2220
300	0.04	3150	2640	2220

# Design Choice for Rare Adverse Outcome

- Clustering matters less with large # of small clusters than with smaller # of larger clusters
- If feasible, recruit large # dental practices and have each one enroll a small number of patients
- Cluster-crossover saves large # patients, if additional time required is acceptable
- For a new material/procedure, step-wedge is attractive because everyone starts with the traditional treatment and then switches to new
- If feasible to alternate treatments, use stratified design to minimize # patients needed

# Example 3: Community-Based Trial of Caries Prevention

- Compare 2 educational interventions for caries prevention
- Outcome = caries incidence at individual level
  - Assume incidence = 25% vs 40%
- If no clustering need 304 patients for 80% power, with type I error 0.05\*
- Assume 20-40 communities available (schools, dental practices, etc.).

# Example: Caries Prevention

# Patients required for power 0.8 to detect 40% vs 25%  
with 20 communities

ICC	Cluster-random.	CC (conserv.)	CC (optimal)
0.01	356	328	304
0.02	428	352	300
0.03	542	384	296
0.04	744	420	292

# Example: Caries Prevention

# Patients required for power 0.8 to detect 25% vs 40%  
with 40 communities

ICC	Cluster-random.	CC (conserv.)	CC (optimal)
0.01	328	320	304
0.02	352	328	304
0.03	384	336	296
0.04	420	344	296

# Design Choice for Community-Based Prevention Study

- Larger # communities does result in reduced number of individual participants, but savings likely more than offset by extra costs of running study in more communities
- Consider total cost of study in terms of per community cost (may be very high) and per participant cost (perhaps not so high)
- Cross-over of communities has problems of carryover effects (lasting behavior changes).

# Summary

- Completely randomized design is useful for single-center trials with no need for stratification or clustering
- Stratified design helps to increase precision (and power) and ensure balanced treatment groups
- Cross-over design increases precision but is susceptible to carry-over effects (especially split-mouth design)
- Cluster-Randomized design prevents contamination but requires larger sample size
- Cluster Crossover design regains some loss of power due to clustering but trial length is increased
- Step-wedge design useful with rapid enrollment and outcome assessment and for new experimental trts
- Optimal design depends on: type of outcome, treatment effect, ICC, IPC, # clusters, risk of contamination, ...