Controversies and Unresolved Issues in the Design of Randomized Controlled Trials Testing Clinical/Behavioral Public Health Interventions

Part III: Purpose and Design of Pilot RCTs. Concepts and Strategies

UCSF CAPS Methods Core Seminar

October 23, 2018

Steve Gregorich

Pilot RCTs

The RCT is the gold standard research design in medical sciences

. Guidance on the role, design, conduct & reporting of full-scale RCTs
    . Countless textbooks
    . World Medical Association Declaration of Helsinki
    . CONSORT
    . etc.

Pilot RCTs are often conducted before launching a full-scale RCT

. No similar guidance about pilot RCTs & their purpose is debated
    . Traditional perspective
    . Alternative perspective

. Funding challenges

Pilot RCTs. The Traditional Perspective

Size

. How many participants should be enrolled?

. If too large, then the study becomes a definitive RCT, not a pilot.

Reviewers tend to view $N \geq 100$ to suggest 'large,' in this context (i.e., $n=50$ per group)

. What defines the lower bound for pilot RCT sample size?

More on this later

. Goldilocks: not to small, not too large, 'just right': $60 \leq N \leq 100$, usually.

Pilot RCTs. Goals of the Traditional Perspective

Logistics

. Feasibility and Acceptability (F&A) of study procedures

Recruitment: Sufficient numbers and pace

Randomization: Participants willing to be randomized

Fidelity: Intervention delivered as intended

Adherence: Participants following study protocol

Assessment: Valid, reliable, acceptable, complete

Retention: Participants complete study

Statistics

. Obtain intervention effect size estimates

Inform power analysis for subsequent full-scale RCT
(this includes preliminary evidence of efficacy)

Pilot RCTs. Goals of the Alternative Perspective

Logistics

. Feasibility and Acceptability (F&A) of study procedures

Recruitment:	Sufficient numbers and pace
Randomization:	Participants willing to be randomized
Fidelity:	Intervention delivered as intended
Adherence:	Participants following study protocol
Assessment:	Valid, reliable, acceptable, complete
Retention:	Participants complete study

~~**Statistics**~~

~~. Obtain intervention effect size estimates~~

~~Inform power analysis for subsequent full-scale RCT~~

~~(this includes preliminary evidence of efficacy)~~

Goal: demonstrate that a full-scale RCT can be conducted successfully.

I.e., full-scale RCT is unlikely to fail because of logistical futility

Kraemer, H.C. et al (2006). Caution regarding the use of pilot studies to guide power calculations for study proposals. *Archives of General Psychiatry*, 63, 484-489.

Pilot RCTs. The Alternative Perspective

Why should we avoid obtaining effect size estimates from pilot RCTs?

. Low precision

. Fallible decision making

. Equipoise & Ethics

Pilot RCTs. The Alternative Perspective

Why should we avoid obtaining effect size estimates from pilot RCTs?

Low precision

Most agree: pilot RCTs are under-powered wrt tests of group differences

Issue: NIH reviewers still want effect size estimates from pilot RCTs (!)
An under-powered study yields unreliable effect size estimates.

Example

. Two-group pilot RCT with *n*=40/group; 80% retention → *n*=32/group

. Expected widths of 95% CI for group difference

. Continuous *Y*: 95% CI width equals 1.0 std dev

. Binary *Y*: If *true* prevalence is 50% in both groups, then
95% CI covers *Intervention* group prevalence from
25% (OR=.33) to 75% (OR=3.0)

Pilot RCTs. The Alternative Perspective

Why should we avoid obtaining effect size estimates from pilot RCTs?

Fallible Decision Making

Kraemer et al took it their argument one step further

If a pilot study overestimates the true effect size…
Larger assumed effect size requires smaller sample to achieve power

So, the sample size chosen for the full-scale RCT will be too small.

I.e., the full-scale RCT will be under-powered.

If a pilot study underestimates the effect size…
Discouraging. Often will not proceed to a full-scale RCT

Result

Too many full-scale RCTs that either are under-powered or not conducted

Pilot RCTs. The Alternative Perspective

Why should we avoid obtaining effect size estimates from pilot RCTs?

Equipoise & Ethics

Equipoise: Honest uncertainty about whether the experimental intervention will provide a benefit relative to the comparator

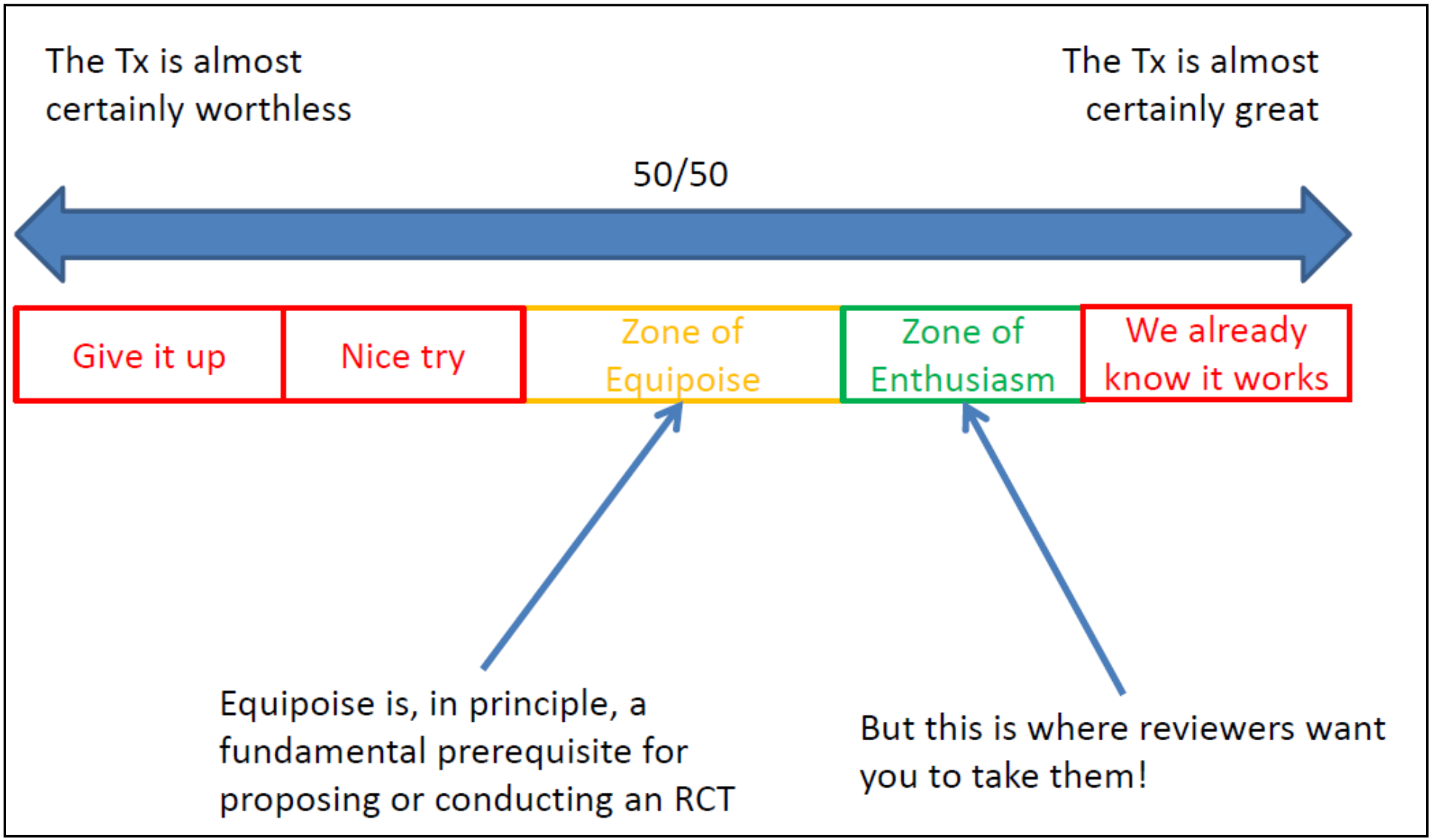
Equipoise provides the ethical basis for randomly assigning participants to different interventions in RCTs

Issue: Reviewers want pilot RCTs to show prelim. evidence of efficacy i.e., show the 'promise' of an experimental intervention

This goes against the concept of equipoise.

Pilot RCTs. The Alternative Perspective

Equipoise and Ethics



Ref: Freedland, K. (2016) *Feasibility and Pilot Studies*. (Slide set) https://www.sbm.org/UserFiles/file/Seminar14_Freedland.pdf

Pilot RCTs. The Alternative Perspective

Why should we avoid obtaining effect size estimates from pilot RCTs?

Risks

. Unreliable effect size estimates

. Full-scale RCTs that are conducted but underpowered or abandoned

. Full-scale RCTs that lack equipoise

Pilot RCT goal summary

. Acceptability & Feasibility
  Demonstrate whether a full-scale RCT can be conducted successfully

. *Not* about a full-scale RCT having a good chance of a signif. outcome.

. *Not* about efficacy, effectiveness, or safety

  Rather that the full-RCT will be a reasonable test of the intervention

Pilot RCTs are about logistics, not statistics

Pilot RCTs. Proposal Strategies

Subject to change: Scientific culture is shifting

. Make the case for a focus on Feasibility & Acceptability

E.g.,

> "Following noted experts and NIH guidance, we acknowledge that pilot RCTs are too small to allow for reliable effect size estimates or sufficiently powered statistical tests and instead should focus on the feasibility and acceptability of a subsequent full-scale RCT [REFS]."

There are several papers to cite. The following are primary.

Kraemer, H.C., Mintz, J., Noda, A., Tinklenberg, J., Yesavage, J.A. (2006). Caution Regarding the Use of Pilot Studies to Guide Power Calculations for Study Proposals. *Archives of General Psychiatry*, 63, 484-489.

NIH/NCCIH. *Pilot Studies: Common Uses and Misuses*. Website: https://nccih.nih.gov/grants/whatnccihfunds/pilot_studies.

NIH/NCCIH guidance was a turning point for me wrt proposing F&A pilot RCTs

Pilot RCTs. Proposal Strategies

Propose specific Feasibility & Acceptability outcomes of a pilot RCT

F&A Construct	Measure	Threshold
Screening	# opting out; # screened by phone per week	No threshold; descriptive
Recruitment	# enrolled per week	Average *X* per week for *Y* weeks
Randomization	Proportion who enroll, complete onboarding; performance of randomization procedures	*X* participants onboarded & randomized by time *Y*
Fidelity of intervention delivery	<*specific to intervention(s)*>	<specific to intervention(s)>
Participant intervention adherence	<*specific to intervention(s)*>	*X*% of INT participants will complete *Y* sessions
Retention	Group-specific retention rates; reasons for dropout	*X*% retention at FU *Y*
Assessment protocol	Duration of battery; proportion completed; participant feedback	*X*% of all subjects complete all assessments
Acceptability to participants (other)	Satisfaction survey; qualitative feedback	*X*% of all subjects satisfied overall

Pilot RCTs. Proposal Strategies

Feasibility & Acceptability outcomes of a pilot RCT

Analysis plan

Descriptive statistics

. Descriptive statistics of F&A outcomes compared to threshold values

. Descriptive statistics of 'clinical' study outcomes

. No inferential statistics—not even 'exploratory' modeling

Sample Size

Chosen based upon subjective judgement

"Because the aim of this pilot RCT is to assess feasibility and acceptability of the research protocol, the sample size of *N*=*XX* (*YY*/group) was set for practical reasons and not driven by hypothesis testing or estimating effect sizes."

If spelled-out, reviewers tend to accept the alternative perspective.
However, some still want precision estimates around F&A thresholds!

Power/sample size calculations for the full-scale RCT

Issue

How to choose an effect size to inform power analysis
when proposing the subsequent, full-scale RCT?

Options

. (Minimum) Clinically Important Differences: (M)CID
AKA, Clinically Meaningful Differences: CMD

CIDs are not currently available for many outcomes.
Defining CIDs can be a lengthy process.

Consider defining CIDs as an Aim of your R34 project, e.g.,
Via stakeholder patient, clinician, practitioner, community,
and/or policymaker groups

. Use 'benchmark' thresholds for effect size (e.g., 'small-medium' effect)
Requires a review group with an 'agreeable' culture.

Resources

Alternative Perspective on Pilot RCTs

Kraemer, H.C., Mintz, J., Noda, A., Tinklenberg, J., Yesavage, J.A. (2006). Caution Regarding the Use of Pilot Studies to Guide Power Calculations for Study Proposals. *Archives of General Psychiatry*, 63, 484-489.

NIH/NCCIH. *Pilot Studies: Common Uses and Misuses*. Website: https://nccih.nih.gov/grants/whatnccihfunds/pilot_studies.

Leon, A.C., Davis, L.L., Kraemer, H.C. (2011). The role and interpretation of pilot studies in clinical research. *J of Psychiatric Research*, 45-626-629.

Freedland, Kenneth. (2016). *Feasibility and Pilot Studies*. Slide set: https://www.sbm.org/UserFiles/file/Seminar14_Freedland.pdf

…more out there…

Resources

Fidelity Monitoring

Borrelli, B. (2011). The Assessment, Monitoring, and Enhancement of Treatment Fidelity in Public Health Clinical Trials. *J Public Health Dentistry*, 71, S52-S63.

Borrelli. B., Sepinwall, D., Ernst, D., Bellg, A.J., Czajkowski, S., Breger, R., DeFrancesco, C., Levesque, C., Sharp, D.L., Ogedegbe, G., Resnick, B., Orwig, D. (2005). A new tool to assess treatment fidelity and evaluation of treatment fidelity across 10 years of health behavior research. *J Consulting & Clinical Psychology*, 73, 852-860.

Bellg, A.J., Borrelli, B., Resnick, B., Hecht, J., Minicucci, D.S., Ory, M., Ogedegbe, G., Orwig. D., Ernst, D., Czajkowski, S., Treatment Fidelity Workgroup of the NIH Behavior Change Consortium (2004). Enhancing treatment fidelity in health behavior change studies: best practices and recommendations from the NIH Behavior Change Consortium. *Health Psychology*, 23, 443-451.

Resnick, B., Bellg, A.J., Borrelli, B., DeFrancesco, C., Breger, R., Hecht, J., Sharp, D.L., Levesque, C., Orwig, D., Ernst, D., Ogedegbe, G., Czajkowski, S. (2005). Examples of implementation and evaluation of treatment fidelity in the BCC studies: Where we are and where we need to go? *Annals of Behavioral Medicine*, 29, Suppl 46-54.

Part III

Pilot RCTs are about logistics, not statistics

END