

The Role of Pilot Feasibility & Acceptability Studies in Randomized Controlled Trials

UCSF Center for Aging in Diverse Communities, Analysis Core

Steve Gregorich, Anita Stewart, and Tor Neilands

A pilot randomized controlled trial (RCT) study helps investigators prepare for a subsequent full-scale RCT. Following the seminal work of Helena Kraemer and co-authors (Kraemer et al, 2006), the emphasis of pilot RCT studies has shifted from being heavily focused on estimating effect sizes to assessing the feasibility and acceptability (F&A) of the study protocol with respect to recruitment, randomization, fidelity of intervention delivery, participant adherence to study protocol and experimental interventions, data collection, and participant retention. Consistent with this new emphasis, the National Center for Complementary and Integrative Health (NCCIH) defines a pilot study as...“a small-scale test of methods and procedures to assess the feasibility/acceptability of an approach to be used in a larger scale study”

(<https://www.nccih.nih.gov/grants/pilot-studies-common-uses-and-misuses>). An important feature of pilot F&A RCTs is that results can be used to identify and inform any needed modifications and to demonstrate whether a subsequent full-scale study employing the resulting protocol can be successfully conducted (Leon, 2011). Thus, pilot F&A RCTs have been defined as those aiming to ...“field-test logistical aspects of the future study and to incorporate these aspects into the study design” (Kistin, 2015).

- Leon AC, Davis LL, Kraemer HC. The role and interpretation of pilot studies in clinical research. *Journal of psychiatric research*. 2011;45(5):626-629.
- Kistin C, Silverstein M. Pilot studies: a critical but potentially misused component of interventional research. *JAMA*. 2015;314(15):1561-2.

Both quotes above—from the NCCIH webpage and the Kistin article—use the word 'test' in an informal way. To be precise, pilot F&A RCTs are not designed to formally *test* research questions or hypotheses. By their nature, pilot study sample sizes are too small to support inferential statistics or reasonably precise effect size estimation. Instead, pilot F&A RCTs are designed to assess the feasibility of successfully completing a subsequent full-scale RCT protocol. That is, pilot F&A RCTs focus on *logistics*, not *statistics*.

This document includes resources for investigators who are proposing, planning, conducting, or reporting upon a pilot F&A RCT: key references in the peer reviewed literature; links to web resources on pilot F&A RCTs including the NCCIH webpage, presentation slide sets, and a webinar; as well as proposal boilerplate.

It is intended to complement the other resources available on the CADC website pertaining to Conducting feasibility and acceptability pilot studies.

Part 1: Web Resources

National Center for Complementary and Integrative Health. NCCIH Research Blog. *Pilot Studies: Common Uses and Misuses*. <https://www.nccih.nih.gov/grants/pilot-studies-common-uses-and-misuses>

Kenneth Freedland, Washington University School of Medicine, slide set: Feasibility and Pilot Studies (2016). https://www.sbm.org/UserFiles/file/Seminar14_Freedland.pdf

Catherine Sarkisian, UCLA David Geffen School of Medicine, YouTube: Methods Seminar: Pilot Studies versus Feasibility Studies to inform Future RCTs (2021). <https://www.youtube.com/watch?v=e860p5aFqps>

Steve Gregorich, UCSF School of Medicine, YouTube: "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 (2021).

- Video: <https://www.youtube.com/watch?v=0ueISF3MYnM>
- Slides: <https://cadc.ucsf.edu/sites/g/files/tkssra881/f/191001%20Pilot%20RCT%20Slides%20for%20CAPS%20Updated.pdf>

Part 2: Proposal Boilerplate

Specific AIM 3: To assess the feasibility and acceptability (F&A) of a research protocol evaluating the ZZ intervention as an approach to improve ZZ. A pilot RCT will assess the F&A of a research protocol comparing patients participating in the ZZ intervention versus a usual care (UC) condition. Eligible participants will complete baseline measures and then be randomized for a total study period of 6 months. ZZ patients will <describe what intervention participation entails>. UC patients will receive <describe what control group participation entails>. Primary outcomes for this pilot study are specific to F&A of the research protocol (see Table 1). Primary outcomes to be assessed for use in the subsequent full-scale RCT will include <list of outcomes>. Following noted experts and NIH guidance, we acknowledge that pilot RCT studies are too small to allow for sufficiently powered statistical tests or reliable effect size estimates and should instead focus on feasibility and acceptability of a full-scale ZZ RCT. (cite references listed below)

F&A Construct	Measure	Threshold
1. Screening	# opting out; # screened by phone per week	No threshold; descriptive
2. Subject recruitment	# enrolled per week	Average X per week for Y weeks
3. Randomization	Proportion and number who enroll, complete onboarding, and start intervention; performance of randomization procedures	X participants randomized by Y time
4. Subject retention	Intervention group specific retention rates; reasons for dropout	X% retention at final follow-up
5. Adherence to ZZ intervention	<explain elements of participant adherence to intervention protocol>	X% of participants in ZZ condition will <list primary adherence criteria>.
6. Intervention fidelity	<describe elements assessed to determine fidelity of intervention delivery>	X% of intervention components delivered on schedule. Y% of intervention components delivered with adequate quality
7. Assessment protocol	Duration of battery; proportion completed; subject feedback	75% of all participants complete all assessments
8. Conditions acceptable to participants?	Satisfaction survey; qualitative feedback	75% of all participants satisfied overall; ZZ intervention rated 3+ of 5

Feasibility and Acceptability (F&A). Our primary goal is to assess F&A of the ZZ intervention and the experimental protocol. Formal tests of clinical outcomes or attempting to obtain precise estimates of effect sizes cannot be statistically justified (Kraemer et al., 2006; Leon et al. 2011, Teresi et al. 2022). Pilot studies, by design, cannot definitively test hypotheses, due to their smaller sample sizes and the frequent design adjustments necessary to maximize recruitment, retention, and quality assessment of outcomes. Effect size estimates are not sufficiently precise given the breadth of the confidence intervals. Nevertheless, the proposed pilot will assess whether a subsequent full-scale RCT modeled after this pilot is logistically feasible and acceptable by systematically gathering important information about the study of ZZ. Important aspects of F&A are operationalized in Table 1. Because pilot RCTs are too small to provide precise estimates of any study outcome—including F&A outcomes—we propose threshold values for each F&A criterion. The primary endpoints for a subsequent full-scale RCT are found in Table X and also will be assessed for F&A in this pilot study. <Table X intentionally omitted from this boilerplate.>

Data Collection. Data elements, instruments, and timeline (See Table X). Clinical outcomes will include <list of outcomes>. Secondary outcomes will include <list of outcomes>. <Essentially, describe the clinical outcomes that will be the assessed in the subsequent full-scale RCT and will be collected as part of this pilot RCT to assess F&A of collecting those measures>.

Statistical Analyses. Because pilot studies are too small to definitively test hypotheses or estimate precise effect sizes (Kraemer et al., 2006; Leon et al. 2011, Teresi et al. 2022), we do not propose any inferential statistics. Primary quantitative analyses will include descriptive statistics of the feasibility and acceptability indicators, comparing each statistic (e.g., % retained) to its tabled threshold (above). Above-threshold findings will suggest a reasonable level of F&A for the corresponding aspects of study procedures. Any sub-threshold finding would suggest that remedial modifications to study procedures and/or design would be required prior to moving forward with a full-scale RCT; qualitative analyses of the exit interviews described below will be instructive under this circumstance. We also will examine descriptive statistics of the primary and secondary clinical outcomes.

Power and Sample Size. As the primary aim of this pilot study is to assess feasibility and acceptability of the research protocol for a future clinical trial, the sample size of $N=80$ (40 in each condition) was set primarily for practical reasons and not driven by hypothesis testing or allowing for precise effect size estimates. Effect sizes used to inform power analysis for a future full-scale RCT will be taken from the literature and based on clinically important differences. *<F&A two-group pilot RCTs have some latitude with respect to planned sample size. For many pilot studies of clinical/community/behavioral interventions, we suggest a number in the range 60-90. Below $N=60$ (30/group) may be viewed as too 'small' whereas above $N=90$ might prompt some reviewers to think that the study is 'larger' than a pilot and may be approaching the size of a definitive trial. Use your judgement.>*

References for the above sample text

- NIH/NCCIH (2017). Pilot Studies: Common Uses and Misuses.
https://nccih.nih.gov/grants/whatsnccihfunds/pilot_studies. Accessed February 14, 2018.
- Kraemer HC, Mintz J, Noda A, Tinklenberg J, Yesavage JA. Caution regarding the use of pilot studies to guide power calculations for study proposals. *Archives of general psychiatry*. 2006;63(5):484-489.
- Leon AC, Davis LL, Kraemer HC. The role and interpretation of pilot studies in clinical research. *Journal of psychiatric research*. 2011;45(5):626-629.
- Moore CG, Carter RE, Nietert PJ, and Stewart PW. Recommendations for planning pilot studies in clinical and translational research. *Clinical and Translational Science*, 2011;4, 332-337.
- Thabane L, Ma J, Chu R, Cheng J, Ismaila A, Rios LP, Robson R, Thabane M, Giangregorio L, Goldsmith CH. A tutorial on pilot studies: the what, why and how. *BMC Medical Research Methodology*, 2010;10:1, 1-10.
- Teresi, J.A., Yu, X., Stewart, A.L., Hays, R.D. Guidelines for Designing and Evaluating Feasibility Pilot Studies. *Medical Care*. 2022;60(1):95-103.

Part 3: Selected Topics in the Peer-reviewed Literature

3A. Defining and reporting pilot feasibility and acceptability RCTs

- Kistin C, Silverstein M. Pilot studies: a critical but potentially misused component of interventional research. *JAMA*. 2015;314(15):1561-2.
- Eldridge SM, Lancaster GA, Campbell MJ, Thabane L, Hopewell S, Coleman CL, et al. Defining feasibility and pilot studies in preparation for randomised controlled trials: development of a conceptual framework. *PLoS One*. 2016;11(3):e0150205.
- Leon AC, Davis LL, Kraemer HC. The role and interpretation of pilot studies in clinical research. *J Psychiatr Res*. 2011;45(5):626-9.
- Teresi, J.A., Yu, X., Stewart, A.L., Hays, R.D. Guidelines for Designing and Evaluating Feasibility Pilot Studies. *Med Care*. 2022;60(1):95-103.

3B. Argument against using pilot studies to guide power calculations

Traditionally, pilot studies were used, in part, to estimate effect sizes to inform the sample size needed for a subsequent full-scale RCT. However, this practice has been challenged because pilot study samples are typically 'small,' leading to 'noisy' estimates that can be overly pessimistic (potentially prompting abandonment of the subsequent full-scale trial) or overly optimistic (potentially prompting an underpowered full-scale trial).

Kraemer HC, Mintz J, Noda A, Tinklenberg J, Yesavage JA. Caution regarding the use of pilot studies to guide power calculations for study proposals. *Arch Gen Psychiatry*. 2006;63(5):484-9.

3C. Choosing effect sizes to inform power analyses of the subsequent full-scale trial

Because pilot RCTs cannot be relied upon to provide reasonably precise effect size estimates, the most promising approach is to choose effect sizes that are clinically meaningful, i.e., design the full-scale RCT to detect a clinically/minimally important difference (CID, MID; e.g., Kraemer et al 2006). In some content areas, CIDs are well established. In some cases, CIDs are not formally established but experts generally agree without requiring a formal process. In other cases, a formalized process may be required to decide what constitutes a CID/MID, e.g., convening an expert panel.

Hays RD, Farivar SS, Liu H. Approaches and recommendations for estimating minimally important differences for health-related quality of life measures. *COPD*. 2005;2(1):63-7.

Keefe RS, Kraemer HC, Epstein RS, Frank E, Haynes G, Laughren TP, et al. Defining a clinically meaningful effect for the design and interpretation of randomized controlled trials. *Innov Clin Neurosci*. 2013;10(5-6 Suppl A):4S-19S.

Kraemer HC, Mintz J, Noda A, Tinklenberg J, Yesavage JA. Caution regarding the use of pilot studies to guide power calculations for study proposals. *Arch Gen Psychiatry*. 2006;63(5):484-9.

McLeod LD, Coon CD, Martin SA, Fehnel SE, Hays RD. Interpreting patient-reported outcome results: US FDA guidance and emerging methods. *Expert Rev Pharmacoecon Outcomes Res*. 2011;11(2):163-9.

Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol*. 2008;61(2):102-9.