Published in final edited form as:

JAMA. 2015 October 20; 314(15): 1561–1562. doi:10.1001/jama.2015.10962.

Pilot Studies: A Critical but Potentially Misused Component of Interventional Research

Caroline Kistin, MD, MSc and

Department of Pediatrics, Boston Medical Center, Boston, Massachusetts; and Boston University School of Medicine, Boston, Massachusetts

Michael Silverstein, MD, MPH

Department of Pediatrics, Boston Medical Center, Boston, Massachusetts; and Boston University School of Medicine, Boston, Massachusetts

Pilot intervention studies, commonly defined as small studies carried out in preparation for larger investigations, ¹ are essential precursors to high-quality clinical trials. Although not all trials are preceded by pilot investigations, when they are, pilot studies inform how subsequent trials are conducted and have an important role in controlling the pipeline of intervention development and dissemination by influencing which large-scale trials are ultimately carried out and which are never brought to fruition. However, pilot studies have often involved reporting inconsistencies, misapplication of research techniques, and misinterpretation of results, with significant implications for intervention research. ^{1–3} These suboptimal research practices likely emanate from the understandable but methodologically unsound view of pilot studies as mechanisms to obtain preliminary answers to primary research questions. ^{1–3} The risks of this approach may include unjustified and potentially misleading conclusions in the short term as well as misinformed decisions about how—and even whether—to proceed to more definitive studies.

Although widely accepted guidelines exist for how to conduct and report definitive clinical trials, ⁴ there is no equivalent set of guidelines for pilot studies, the objectives of which are more variable. ⁵ Within an increasing scholarly literature on pilot studies, consensus has emerged that pilot studies should focus on feasibility, process, and description, as opposed to group-to-group comparison of outcomes. ^{1,2,5} Despite this consensus, a 2010review demonstrated that 21 of 26 (81%) published pilot studies of interventions incorporated hypothesis testing with sample sizes known to be insufficient. ¹ Instead, pilot intervention studies are most effective when they focus on testing study logistics under the circumstances mimicking the eventual definitive study (ie, field testing), optimizing intervention delivery,

Corresponding Author: Caroline Kistin, MD, Department of Pediatrics, Boston Medical Center, 88 E Newton St, Vose Hall, Third Floor, Boston, MA 02118 (caroline.kistin@bmc.org).

Conflict of Interest Disclosures: The authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest and none were reported.

Role of the Funder/Sponsor: The NIH had no role in the preparation, review, or approval of the manuscript or the decision to submit the manuscript for publication.

Additional Contributions: We thank Thomas Koepsell, MD, MPH (University of Washington), for his insight and guidance in this area. Dr Koepsell received no compensation for his contributions. We also thank the members of the Department of General Pediatrics at Boston University School of Medicine/Boston Medical Center for their thoughtful feedback.

Kistin and Silverstein Page 2

understanding the barriers to and facilitators of eventual dissemination and implementation of an intervention, and obtaining empirical evidence of study parameters to help design the definitive clinical trial.

A definitive randomized clinical trial (RCT) eliminates unmeasured confounding and therefore is the best study design to safeguard internal validity for making inferences about efficacy. However, well-planned RCTs represent a complex synthesis of many components, each of which may be field tested and improved during a pilot investigation. During a definitive RCT, an efficient recruitment strategy, often based on pilot work, typically brings potential research participants in contact with the study team. The study team must explain the trial, including the concept of randomization in understandable terms; determine eligibility according to prespecified criteria; conduct a baseline assessment; and enroll participants, assigning them to the appropriate, randomly determined, interventional path. Study participants are then exposed to the study interventions or conditions—ideally, with fidelity to what investigators had originally planned and potentially proved feasible within the context of a pilot study. To the extent possible, and within ethical boundaries, those participating in clinical trials should not know exactly what they are being exposed to, and the individuals assessing their outcomes also should be unaware of the group to which the study participant was assigned. If each of these study components works as intended during the definitive RCT, with a limited number of unanticipated safety problems, postrandomization exclusions, or protocol changes, a statistically valid comparison between study conditions can be made, and internally valid conclusions can be drawn.

However, internal validity can be significantly compromised by insufficient piloting of key study logistics. If, for example, recruitment lags or large numbers of eligible study participants refuse to be enrolled and randomized, sample size goals may not be met, and the study becomes subject to type II error (false-negative results). If eligibility criteria fail to account for an unanticipated phenomenon (eg, presumptive diagnoses being overturned more frequently than projected), there may be a large number of postrandomization exclusions. If follow-up assessors are inadvertently unmasked, outcome estimates may be biased. If there is differential attrition according to study group or clinical status, the effect of an intervention can be overestimated or underestimated. Well-designed pilot studies are investigators' best defense against these, and other, threats to validity. Thus, the principal goal of a pilot study is to field-test logistical aspects of the future study and to incorporate these aspects into the study design.

Another goal of conducting a pilot study is to optimize intervention delivery, with specific attention to adherence and fidelity. In many cases, a pilot study can identify problems with adherence and use qualitative research or quality improvement techniques to understand and correct them. If suboptimal adherence is modifiable and is not recognized and corrected in a pilot study, a subsequent trial is vulnerable to a number of threats to validity. First, a proper intention to treat analysis may be "diluted" if a substantial number of individuals in the intervention group never receive the intervention or cross over to receive the comparator intervention. Thus, a potentially efficacious intervention—delivered poorly—will appear to have no effect. Second, if the traditional trajectory of interventional research is to move from pilot studies to efficacy studies, to effectiveness studies, and then to dissemination, a poorly

Kistin and Silverstein Page 3

understood or insufficiently field-tested intervention delivery system (for example, one that relies on already overstretched primary care clinicians, or one for which the fidelity cannot be maintained in a certain context) may prematurely halt the trajectory of an intervention along this continuum. Pilot studies—particularly those with a preconceived adaptive trial design—thus afford investigators the opportunity to tweak intervention delivery in real time, as if the pilot study were functioning as a quality improvement project. In such adaptive pilot studies, the intervention offered to the first enrolled study participant is by definition different from that offered to subsequent participants.

Pilot studies also present an opportunity to begin to understand the barriers and facilitators to eventual intervention dissemination and implementation. Techniques such as failure modes and effect analyses, root cause analyses, and ethnographic interviewing are well known to other disciplines and have recently begun to be discussed in the health services and clinical research literature and incorporated into research practice. Such techniques represent efficient ways to understand how and why implementation of a clinical innovation—or elements thereof—can fail and what the consequences of such failures are. In pilot studies, identifying such failure modes can lead to their correction or at least inform the design of the subsequent clinical trial to minimize problems with recruitment, intervention delivery, and follow-up.

Another goal of pilot studies is to obtain empirical estimates of statistical parameters to inform power calculations, and other design effects, of a subsequent trial. A common approach is to treat the pilot study as a mechanism to obtain an estimate of effect size—which is then often used to calculate sample size for the definitive trial. An article by Kraemer et al³ illustrates the potential problems with this approach. Specifically, in a small pilot study, an effect size estimate is more likely than not to either overestimate or underestimate the true effect size. If the true effect size is underestimated, a potentially worthwhile main study may be terminated early or reviewed unfavorably on the basis that the pilot study failed to demonstrate clinically significant differences. If the true effect size is overestimated but still used to inform sample size calculations, the clinical trial will be underpowered. Caution should thus be used when reporting the results of pilot investigations.

Pilot studies also represent an opportunity to estimate other critical study parameters. These parameters might include the within-group standard deviation of continuous outcome measures, proportion of study participants who experience a key outcome event spontaneously without intervention, or level of correlation over time for a repeated outcome measure. For multicenter studies, a pilot study of at least 2 sites might help estimate the influence of center effects on sample size requirements. For studies in which study participants are clustered within intervention clinicians or centers, the extent of this clustering effect is important to understand subsequent power calculations and other potentially important design effects.

Pilot studies have an important role in intervention development and, when interpreted appropriately, can increase the efficiency and validity of subsequent clinical trials by improving subject recruitment and intervention delivery and by enabling investigators to

Kistin and Silverstein Page 4

more accurately determine sample size requirements.² Too often, pilot studies are instead primarily used to evaluate the preliminary effects of an intervention, which can lead to overestimations or underestimations of the rue effects of an intervention. Such inaccurate estimations can lead to misguided practice or to inappropriate decisions about how to proceed on a trajectory of research. Formal guidelines for the conduct and reporting of pilot studies, similar to those available for definitive clinical trials, could potentially help standardize their use and increase their utility.

Acknowledgments

Funding/Support: This work was supported by National Institutes of Health (NIH) grants 1K23HD078503-01A1 (Dr Kistin) and 1K24HD081057-01 (Dr Silverstein).

References

- 1. Arain M, Campbell MJ, Cooper CL, Lancaster GA. What is a pilot or feasibility study? a review of current practice and editorial policy. BMC Med Res Methodol. 2010; 10:67. [PubMed: 20637084]
- 2. Lancaster GA, Dodd S, Williamson PR. Design and analysis of pilot studies: recommendations for good practice. J Eval Clin Pract. 2004; 10(2):307–12. [PubMed: 15189396]
- 3. 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):484–489. [PubMed: 16651505]
- 4. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ. 2010; 340:c332. [PubMed: 20332509]
- 5. Koepsell, TD.; Weiss, NS. Epidemiologic Methods. New York, NY: Oxford University Press; 2003.
- Langley, G.; Nolan, K.; Nolan, T.; Norman, C.; Provost, L. The Improvement Guide: A Practical Approach to Enhancing Organizational Performance. San Francisco, CA: Jossey-Bass; 1996.
- Meurer WJ, Lewis RJ, Berry DA. Adaptive clinical trials: a partial remedy for the therapeutic misconception? JAMA. 2012; 307(22):2377–2378. [PubMed: 22692168]
- 8. Lewin S, Glenton C, Oxman AD. Use of qualitative methods alongside randomised controlled trials of complex healthcare interventions: methodological study. BMJ. 2009; 339:b3496. [PubMed: 19744976]