

Recommendations for Planning Pilot Studies in Clinical and Translational Research

Charity G. Moore, Ph.D.¹, Rickey E. Carter, Ph.D.², Paul J. Nietert³, and Paul W. Stewart, Ph.D.⁴

Abstract

Advances in clinical and translation science are facilitated by building on prior knowledge gained through experimentation and observation. In the context of drug development, preclinical studies are followed by a progression of phase I through phase IV clinical trials. At each step, the study design and statistical strategies are framed around research questions that are prerequisites for the next phase. In other types of biomedical research, pilot studies are used for gathering preliminary support for the next research step. However, the phrase “*pilot study*” is liberally applied to projects with little or no funding, characteristic of studies with poorly developed research proposals, and usually conducted with no detailed thought of the subsequent study. In this article, we present a rigorous definition of a pilot study, offer recommendations for the design, analysis and sample size justification of pilot studies in clinical and translational research, and emphasize the important role that well-designed pilot studies play in the advancement of science and scientific careers. Clin Trans Sci 2011; Volume 4: 332–337

Keywords: pilot studies, pilot study design, sample size, power calculations, confidence intervals

Introduction

The use of the phrase “*pilot study*” in clinical and translation research is widespread. A MedLine search of the term in the title of articles published during the past year alone (from January 2010 through March 2011), for example, yielded over 19,000 publications. Through our own work as members of biostatistics, epidemiology, and research design (BERD) resource units in institutions that have received Clinical and Translational Science Awards (CTSAs) from the National Institutes of Health, we have found that a high percentage of investigators seeking consultations have said they are either conducting or planning to conduct pilot studies. The purpose of our article is to provide a clear definition of a pilot study in the context of clinical and translational science and to offer recommendations about appropriate pilot study design. To achieve this goal, we first present some commonly observed scenarios that are viewed as problematic or inappropriate application of the term “pilot” to describe the study. The paper concludes with the recommendations.

Common problematic scenarios

We define “pilot studies” as preparatory studies designed to test the performance characteristics and capabilities of study designs, measures, procedures, recruitment criteria, and operational strategies that are under consideration for use in a subsequent, often larger, study. Although pilot studies are often considered less important than other studies, they require no less care in planning—particularly if they are to be published. Successful pilot studies provide optimal information needed to prepare for future studies and thereby help to move clinical and translational investigative careers forward. There are several common scenarios in which misuse of the phrase *pilot study* creates pitfalls.

First, the label “pilot study” is liberally applied to projects with little or no funding. This misuse stems from a self-perpetuating misconception that pilot status and lack of funding are mutually justifying. This view fails to recognize that the study’s design and scientific integrity are usually compromised by inadequate funding. The study is undermined when study coordination, technical expertise, database management, and biostatistical

assistance are sacrificed. The sample size is often matched to the level of funding without regard to the likelihood of achieving the specific aims. As Perry emphasized, “Calling a study a pilot is no excuse for a small sample size or for inadequate methodological rigor.”¹

Second, the “pilot” label is often seen on vague, poorly developed research proposals. Common manifestations of this pitfall are vague research questions and poor alignment of the stated specific aims with the plans for statistical analyses and sample size justification. For example, a misalignment has occurred when the stated aims are to evaluate recruitment feasibility and treatment tolerability, but the proposed analysis plan and sample size justification focus on a hypothesis test of treatment efficacy. In some cases many outcomes (efficacy, surrogates, biomarkers) are described, and thus the study may be better described as exploratory, where the goal is hypothesis generation or refining existing hypotheses rather than testing feasibility. These types of studies are particularly valuable for giving direction to future investigations. For the purposes of this manuscript we consider exploratory studies to be different than pilot studies.

Third, in many cases, the apparent rationale for using the *pilot* label is simply that the study precedes a more costly study. These pilot study proposals do not provide any details about the subsequent study other than “the data from this pilot will inform a future larger R01-funded study.” The pitfall is that there is no assurance that the pilot study will provide the particular information needed to inform the successful design and execution of the subsequent study. The overarching research question and potential design of subsequent studies are critical elements in designing and justifying any proposed pilot study.

Finally, as CTSA BERD faculty members, we have observed that many junior clinical and translational investigators become involved in a nonproductive research strategy that involves designing and implementing a series of pilot studies that yield “negative” results ($p > 0.05$) and are subsequently replaced by negative results from additional pilot studies (Figure 1). The investigators do not take the time to publish what they have

¹University of Pittsburgh, Medicine, Pittsburgh, Pennsylvania, USA; ²Mayo Clinic, Health Sciences Research, Rochester, Minnesota, USA; ³Medical University of South Carolina, Medicine, Charleston, South Carolina, USA; ⁴University of North Carolina, Biostatistics, Chapel Hill, North Carolina, USA.

Correspondence: C Moore (moorecg@upmc.edu)

DOI: 10.1111/j.1752-8062.2011.00347.x

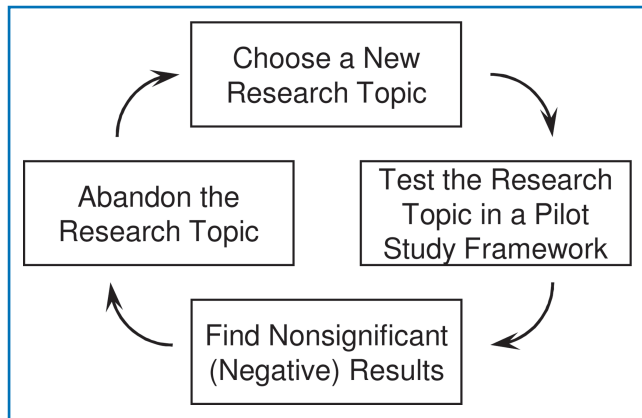


Figure 1. Nonproductive scientific strategy involving the use of pilot studies.

learned from the pilot study because “they didn’t find anything.” This endless cycle of poor staging of pilot studies without setting goals to move forward uses a substantial amount of the investigator’s time and resources and fails to advance scientific understanding and the investigator’s career.

Recommendations for investigators considering a pilot study

Here we provide recommendations concerning the preparatory relationship of the pilot study to one or more subsequent studies, protocol design and analysis for the pilot study, sample size considerations for the pilot and the future study, and the role of pilot studies in career development. We recommend that pilot study research proposals are best developed via the organization of a master protocol document which is updated throughout the course of the research and usually has sections specifying objectives, study design, enrollment and withdrawal, schedule, procedures and evaluations, safety and clinical monitoring, biostatistical considerations, database management plans, and quality assurance, and appendices of data collection instruments. A working document such as this serves as the basis for grant proposals and publications.

1. Keeping the Next Study in Mind

Clinical and translational research spans a spectrum of efforts ranging from work at the molecular level to widespread testing of a treatment in a community setting. “Stages” of testing in drug development are highly accepted and implemented for drug development starting with preclinical studies followed by a progression of phase I through phase III clinical trials testing safety, efficacy, and effectiveness. Some of these trials are deemed “pivotal” and may be immediately preceded by a preparatory pilot study. At each step, the study design and statistical strategies are framed around research questions that are prerequisites for the next phase, but the overarching goal is testing the new drug for effectiveness in the targeted population.

In contrast to drug development, the “staging” or “sequencing” of studies is often not as well defined in when the research questions focus on nondrug intervention development or other areas where interventions are not being developed. We have observed in these settings that not enough attention is given to staging subsequent research studies. Investigators who are proposing a pilot study must have an understanding of the overarching research question and the stage of scientific evidence building that their study

addresses.^{1,2} Furthermore, they must understand whether the main purpose of their pilot study is to investigate a process (e.g., recruitment feasibility, protocol refinement, willingness of physician to randomize or patient to be randomized) or an outcome (e.g., complication rates or efficacy measures).

The aims and methods of a pilot study must be aligned with the goals of the subsequent study. For example, in the field of anticancer drug development, Schoenfeld³ defined pilot studies as small, nonrandomized clinical trials with three major functions—developing or refining the protocol for a new treatment, detecting unacceptable complication rates, and estimating potential efficacy relative to historical controls—all for the sole purpose of planning subsequent phase 3 studies. Dobkin⁴ has argued for three stages of pilot studies to inform and strengthen implementation of a multicenter randomized trial. While Dobkin’s recommendations were focused on developing motor interventions for disabled patients, the recommendations are broadly applicable and mirror the traditional drug development paradigm. The three stages correspond to consideration-of-concept studies, development-of-concept studies, and demonstration-of-concept pilot studies. This staging is advocated not only for a single study team but for groups of study teams working in the same area of intervention development.

In general, pilot studies contribute to the development and design of future, more costly, primary studies by clarifying and sharpening the research hypotheses to be studied, identifying relevant factors that could create barriers to subsequent study completion, evaluating the acceptability of methods and instruments to participants, measuring the time required for study participation, and providing concrete estimates of the expected rates of missing data and participant attrition.⁵

2. Design the Pilot Study with Carefully Specified Aims and Rationale

Objectives and aims

In practical research, the aims of small studies are often a mix of some of the following: pilot efforts, exploratory efforts, and definitive evaluations. For clarity in this report we consider only studies whose aims all concern “pilot study” objectives.

Pilot study aims should serve to identify and address issues that could occur with respect to future study conceptualization, study design, sample size, sample selection, data collection, data management, and data analysis.^{6,7} The aims of pilot studies can range from evaluating the feasibility of protocol implementation to investigating the potential mechanisms of efficacy for a new intervention.⁵⁻⁹

Once the aims are clearly defined, the specification of the primary outcome variables should follow. In some cases, the investigators will seek to generate preliminary data on an established outcome of interest. But if the investigators seek to validate feasibility or logistics, they must consider how to measure these constructs. Often, a feasibility or logistical aim can be distilled into a proportion. For example, if the time from disease onset to treatment is critical, as is the case with stroke treatment, a pilot study could quantify the number and proportion of patients who can undergo the steps of recruitment, informed consent, enrollment, and randomization within the time frame of interest. If the proportion is not high (e.g., >80%), the investigators will need to modify these steps to ensure study feasibility.

Study design

It is vital that the design of the pilot study be guided by the aims of the pilot study. Although virtually any type of design could be used, it is often advantageous for the design of the pilot study to be similar to that of the larger subsequent study, especially if the aim is to assess feasibility. It is also advantageous for the pilot study to recruit participants from the same study population that will be used for the subsequent study.⁵ For example, if one of the aims of the pilot study is to obtain statistical estimates of interparticipant and intraparticipant variance parameters for a clinical outcome, the pilot study will require a study design that incorporates repeated measures and a group of participants drawn from the same population to be used in the subsequent study.

The following set of questions should be answered as a part of the design of the pilot study: (1) What will be the study design of the subsequent study? (2) What statistical strategies and methods will be used in the subsequent study? (3) What novel or untested information is needed to plan the subsequent study, and when will this information be needed? (4) Is the proposed design of the pilot study adequate to obtain the information? (5) What information cannot be obtained, and how will the lack of this information be handled? (6) What generalizable knowledge can be used to plan the pilot study or subsequent study?

Statistical analysis plan

Regardless of the type of study, the hallmark of a well-developed analysis plan is a carefully constructed strategy for achieving each aim. An outline of statistical methods for each aim can serve as a summary, but the details should be fully developed in written paragraphs. Analyses for pilot studies should mainly rely on estimation (point and interval estimation) and involve only limited hypothesis testing within the scope of the original aims.

In a pilot study, the aims should focus on endpoints other than efficacy and safety measurements. For example, they should focus on feasibility. In many cases, quantitative measures of feasibility are not employed. If the aims are well-defined in terms of useful measures, the statistical consideration section of a protocol can follow traditional outlines with details pertaining to endpoint definition, sample size justification, planned analyses, and pilot-specific considerations. These include the justification of the error rate selection,³ accrual expectations, and plans to use the information to guide the design of the subsequent study. While these plans should be emphasized throughout the study protocol, the statistical considerations should provide a clear rationale to support the next steps in the study process, including “go versus no go” decisions.

Cook et al. described a well-designed pilot study for the prophylaxis of thromboembolism. They judiciously chose to study 120 subjects to refine their protocols and maximize the chances of identifying problems that may arise in the larger multicenter trial. Their analyses focused simply on estimating protocol adherence and recruitment rates, comparing observed estimates to ones that were specified *a priori*.¹⁰

3. Justify the Choice of Sample Size

It is not unusual for study proposal reviewers to encounter a statement such as the following from an investigator: “No sample size justification is needed because of the pilot nature of the proposed study.” On the contrary, proposals for pilot studies are not exempt from the need to have a clear and well-reasoned rationale for the number of participants to be included.

Alignment of the rationale with the scientific aims of the research is critical.

Choosing a sample size for pilot studies requires judgment and aims-specific considerations of the issues of practical feasibility, the details of the statistical analysis plans, and the research risks, including the risks that the study aims will not be achieved. Small samples may be appropriate for aims such as pilot-testing a data management system, demonstrating the ability to execute a specific research protocol, or testing acceptability and adherence to a new online disease management intervention. If the aim is to evaluate the feasibility of a protocol across multiple clinical sites, the study could require a small sample for each site but a relatively large total sample. In addition, it is important to keep in mind that some statistical methods are valid only for large samples. Writing a clear and compelling justification for the proposed sample size therefore requires an intellectual analysis of the expected benefits, risks, and costs of the study and the knowledge of which statistical methods are optimal for the study. “Value of information” techniques are becoming more recognized as alternatives to traditional sample size estimation methods that often rely on imprecise parameter inputs.¹¹

Power calculations for hypothesis testing

Some pilot studies will require hypothesis testing to guide decisions about whether further research developments are needed before the larger subsequent study can be undertaken. For example, statistical hypothesis testing is appropriate for addressing research questions such as the following: Is the RNA assay more accurate and more precise than the antigen assay? Is the taste of a particular dietary supplement acceptable to at least 95% of the target population? In cases such as this, the power of the hypothesis test will depend on the choice of sample size and will be a function of hypothesized parameter values. It will therefore be useful for the investigators to calculate the power when different sample sizes are used. For example, sample sizes of 30, 50, and 70 provide 48%, 78%, and 84% power to detect an acceptance rate of 85% or lower if the true acceptance rate in the population is 95% using a one-sided binomial test of size $\alpha = 0.05$.

The literature on hypothesis testing provides formulae for power calculations in approaches that range from the simplest single-sample study to the most complex multilevel and multigroup longitudinal clinical trial. At a minimum, any approach requires specification of the alpha (type I, false-positive) error rate to be considered, the smallest difference in the distribution of the outcome values or event rates to be considered meaningful (clinically important), and the expected variability in outcome values or event rates.

Schoenfeld³ suggested that preliminary hypothesis testing for efficacy could be conducted with a high type I error rate (false positive rate up to 0.25) and that subsequent testing would provide a control over the chances of a false discovery in spite of the initial elevated error rate. He suggested that this approach could ensure that the aims of the pilot study are completed in a timely manner due to the use of a smaller sample size. If such an approach is taken, it is important to remember that a pilot study should not be designed to provide definitive evidence for a treatment effect. It should, on the contrary, be designed to inform the design of a larger more comprehensive study and to provide guidance as to whether the larger study should be conducted.

A pilot study should be an important step within the context of a line of research that is hypothesis driven. This does not mean, however, that every pilot study must test a hypothesis. The investigators who propose pilot studies and the committee members who provide scientific reviews of the proposals often mistakenly presume that a power calculation is the appropriate criterion for choosing the sample size. But if the aims of the pilot study do not include any hypothesis tests, the power level of a test procedure is not a valid consideration for sample size. Whether or not the study includes a hypothesis test, what is always a valid consideration for sample size is the anticipated level of precision of statistical estimators (see *Estimates, Confidence Intervals, and Anticipated Levels of Precision* below).

Because some scientific review committees demand that power calculations be included in all study proposals, the investigators often feel compelled to include hypothesis tests that are not the real reason for conducting their pilot study. For example, the investigators may feel compelled to add a statistical test of efficacy in a pilot study that is actually intended to assess feasibility. In some cases, they compound the problem by providing power levels that are based on overly simplistic computations and are therefore difficult to interpret. In other cases, they summarize the power analysis and include misleading phrases, such as “The study has 80% power if the treatment difference is...” when in fact they should make it clear that the test procedure in question can be expected to detect a difference with high probability only if the magnitude of the treatment effect is unbelievably large in the target population. In other words, it is believed that the test has low power for differences that are smaller but yet clinically meaningful.

If investigators and committee members are focusing on expectations for a hypothesis test that is unimportant in the pilot study, they will lose sight of the contributions that the pilot study's main aims are designed to make. Again, the primary point is that the sample size justification should be based on considerations, calculations, and analyses that directly align with primary goals of the pilot study.⁹

Estimates, confidence intervals, and anticipated levels of precision

Pilot studies are often used to generate preliminary data to support funding applications or protocol development activities. While there are limitations to this approach, a well-designed pilot study can provide a statistical point estimate (sample estimate) of the population parameter (the true, unobservable value in the target population of patients). For example, the percentage of pilot study participants who tolerate a new treatment or the proportion of eligible patients who provide informed consent for participation would be representative of the tolerance or participation rates if all potential participants were studied. Providing a means to estimate these quantities enables early indicators of treatment tolerability and enrollment feasibility. However, there is statistical uncertainty that must be taken into consideration when the estimates are used to generalize the pilot study findings.

Confidence intervals (CI) represent a plausible range for the true population value by accounting for distributional assumptions. Thus, the use of confidence intervals can play a major role in the sample size justification and analysis plans for pilot studies. Of course, the utility of a confidence interval is directly related to the range of the interval, since the range quantifies the imprecision of a sample estimate. If a confidence

interval spans a large range of values, this means that there are many plausible values for the population parameter. In other words, the estimate is relatively imprecise.

While larger sample sizes will yield a more precise estimate, there is usually a nonlinear relationship between the CI width and the sample size (N). In the case of CI for population means or proportions, the width of the CI is proportional to the square root of $1/N$. As N is increased from 5 to 10, the width of the CI will decrease sharply; however, as N is increased from 10 to 15, a slightly smaller decrease in width will be observed. Eventually, further increases in N have diminishing benefit on precision and those decreasing benefits become outweighed by practicalities such as time for recruitment and funding limitations. van Belle¹² and Julious¹³ discussed cases in which increasing the sample size to 12 participants made a profound difference in the width of confidence intervals for mean response, whereas increasing the sample size beyond 12 participants did not. They referred to this result as a “rule of 12” for continuous variables. We recommend at least 12 participants for pilot studies with primary focus of estimating average values and variability for planning larger subsequent studies. This size is quite practical for most early-stage investigators to conduct within single centers while still providing valuable preliminary information.

In other cases, the aims of a pilot study are defined in terms of a binary outcome such as compliance with a protocol or occurrence of an adverse event. For planning these types of studies, anticipated confidence interval widths can be considered for potential sample sizes. For example, if the observed compliance rate in a pilot study was 80%, then the observed exact confidence interval width would be 30% if $n = 30$, or 20% if $n = 70$. Confidence intervals play an important role when all or nearly all of the participants satisfy (or do not satisfy) a criterion such as compliance or intolerance. Confidence intervals based on binomial theory are able to convey the realistic level of uncertainty about such point estimates when planning these studies.¹⁴⁻¹⁶ For example, if $n = 5$ participants are observed to experience 0 events, the 90% exact upper limit for the event rate would be 37%, i.e., CI is [0%, 37%]. If $n = 10$ participants experience 0 events, the CI is [0%, 21%]. For $n = 20$, the 90% CI is [0%, 11%]. The investigator must decide what level of precision is acceptable based on consideration of the stage of investigation (evidence needed), risk to the participant, time for study completion, and cost of the study. Better precision in the pilot study will provide greater confidence when planning a subsequent study.

Additional considerations

In the simplest setting, sample size analyses for a two group clinical trial with a continuous outcome require inputs for Type I and Type II error rates, hypothesized means for each group at the end of the trial, and the standard deviation(s) of the outcome. A good use of a pilot study is to obtain estimates of control group means (μ_c) and group-specific outcome variability for sample size analyses, including calculations of anticipated levels of precision and perhaps power, to be included in larger subsequent studies. The point estimates and confidence intervals for the population control group mean and standard deviations (σ) should be calculated to identify plausible values of μ_c and σ . These estimates and their confidence intervals should be combined with estimates from existing literature to provide a range of possible parameter values. Browne¹⁷ recommends using an upper one-sided confidence limit

($\hat{\sigma}_{UCL}$) of at least 80% for standard deviation inputs from pilot studies. As Browne¹⁷ noted, if the investigator “simply uses the [observed] sample standard deviation from a small pilot sample, the chances of actually achieving the planned power may be as low as 40%.” Browne’s approach takes into account the stochastic nature of the data from the pilot study as well as the stochastic nature of the data from the subsequent study.

As stated previously, the sample size analyses for a two group clinical trial with a continuous outcome require inputs for hypothesized means for each group at the end of the trial. The input for the intervention group mean should be the clinically meaningful value that is truly hypothesized at the end of the study if the intervention is effective.^{18–21} The group means can be used to calculate an effect size by taking their difference and dividing by the standard deviation of the outcome. In the special situation of pilot clinical trials, the power analysis for that larger trial should not be based on the estimated effect ($\hat{\mu}$, the observed mean difference) or the effect size from the pilot study.²² If the estimate was large in the pilot study, using it would result in largely underpowered trials. Also, the decision to move forward to a larger trial should not solely be based on the estimated effect size from the pilot study. If the estimate was small in the pilot study, using it would result in a high likelihood of not investigating truly efficacious interventions.²² Because pilot clinical trials are usually small, they have a high likelihood of imbalances at baseline, leading to unreliable estimates of treatment effects.⁹ And because study populations in pilot clinical trials are typically enrolled within a single study site, the participant population tends to be more homogeneous than that typically enrolled in larger trials and thus contributes to the optimistic overestimation of treatment effects.

Multiple pilot studies may be conducted prior to a larger multicenter clinical trial which test feasibility of the intervention and short-term mechanistic outcomes. Using this approach, nonrandomized or randomized pilot studies of feasibility for the intervention(s) should be conducted prior to a randomized controlled pilot clinical trial. Much can be learned from the feasibility study that may change the population, intervention, protocol, outcomes, etc., possibly requiring an additional pilot study if a substantial modifications are made. Once the intervention is finalized and well defined from the pilot studies, a mechanistic pilot randomized clinical trial (proof-of-principle) may be planned with the targeted mechanism as the primary outcome and sample size analyses based on that outcome. “Pilot studies” that propose a test for potential efficacy for patient relevant outcomes should be designed as Phase II studies using the single-arm or multiple-arm study design and methodology that best answers the research question.^{23,24} These studies are often planned and conducted using higher levels of type of I rates (10–25%) which are consistent with “screening for potential efficacious treatments” to be tested subsequently in larger trials. Type II error rates should be kept lower ($\beta \leq 10\%$) to protect against not putting forward beneficial interventions for additional testing.

4. Pilot Studies and Career Trajectories

At institutions that have a CTSA from the National Institutes of Health, funding is available for clinical and translational pilot studies. These studies are intended to allow investigators to obtain preliminary data and refine procedures and hypotheses to develop protocols for more definitive subsequent studies and

to generate grant proposals for funding from national agencies. In this context, the investigators must be able to propose a productive trajectory for their research agenda, which means that they must have the protocol for the future study in mind when they are planning the initial study.

For junior and developing investigators, the benefits of pilot-funded studies are numerous. A pilot study can provide an ideal opportunity to gain experience with new procedures, new collaborations, protocol writing, funding applications, and institutional review board applications. The preliminary data from a pilot study is critical for convincing funding bodies that the science proposed is feasible and that the study investigator is competent. In this regard, the scientific rigor demonstrated during preliminary investigations often serves as a surrogate for scientific rigor in a more confirmatory setting.

Pilot-funded studies can provide a time-efficient mechanism for professional development but only if their scientific rigor is not compromised. A danger here is that a long series of pilot studies without the subsequent confirmatory studies may be detrimental to an investigator’s career, especially if there is not a clear research theme. To avoid this problem, investigators should be encouraged to publish their results from pilot studies with absolute transparency of the purpose of the study.²⁵ Regardless of whether the pilot studies yield positive or negative results, they can provide valuable information for future research with respect to design, methods, and instruments.⁵ Unfortunately, many investigators make little or no effort to publish negative or unsuccessful findings, and this contributes to publication bias.²⁶

When authors are preparing their work for publication, they should identify their work as a pilot study throughout the manuscript, including in the title, research questions, methodology, and interpretations.¹ It is worth emphasizing here that the goal of publishing results from pilot studies is not to focus on the statistically significant findings but, rather, to provide estimated effects (point estimates with confidence intervals) for all measures of interest and to describe the lessons that have been learned and will be informative in planning subsequent studies.

Summary

Pilot studies are preparatory investigations that provide specific information needed for planning subsequent studies. In many cases, they are designed to test the performance characteristics and capabilities of study designs, measures, procedures, recruitment criteria, and operational strategies that are under consideration for use in a larger subsequent study. They provide the means to evaluate the technical aspects of novel approaches while serving as a platform to generate preliminary data and foster investigator development. To make sure that the information gained through pilot studies can be put to optimal use in a series of studies that build on scientific evidence, investigators must be extremely clear about the purpose of each pilot studies they are planning and must always keep in mind the overarching research question that they will be addressing in their future studies. Although we provided general information about sample size justification using power calculations and precision, we strongly encourage investigators to work with a biostatistician in developing the aims, study design, analysis plan, and sample size justification since each study has its unique aspects that need to be considered.

Acknowledgments

This publication was supported in part by grants from the National Institutes of Health (UL1 RR024150, UL1 RR024153, UL1 RR025747, and UL1 RR029882).

References

1. Perry SE. Appropriate use of pilot studies. *J Nurs Scholarsh*. 2001; 33(2): 107.
2. Polit D, Hungler B. *Nursing Research Principles and Methods*. Philadelphia, PA: Lippincott, 1999.
3. Schoenfeld D. Statistical considerations for pilot studies. *Int J Radiat Oncol Biol Phys*. 1980; 6(3): 371–374.
4. Dobkin BH. Progressive staging of pilot studies to improve phase III trials for motor interventions. *Neurorehabil Neural Repair*. 2009; 23(3): 197–206.
5. Hinds PS, Gattuso JS. From pilot work to a major study in cancer nursing research. *Cancer Nurs*. 1991; 14(3): 132–135.
6. Prescott P, Kl S. The potential uses of pilot work. *Nurs Res*. 1989; 38: 60–62.
7. Jairath N, Hogerney M, Parsons C. The role of the pilot study: a case illustration from cardiac nursing research. *Appl Nurs Res*. 2000; 13(2): 92–96.
8. Hundley V, van Teijlingen E. The role of pilot studies in midwifery research. *Midwives*. 2002; 5(11): 372–374.
9. Arnold DM, Burns KE, Adhikari NK, Kho ME, Meade MO, Cook DJ. The design and interpretation of pilot trials in clinical research in critical care. *Crit Care Med*. 2009; 37(1 Suppl): S69–74.
10. Cook DJ, Rocker G, Meade M, et al. Prophylaxis of Thromboembolism in Critical Care (PROTECT) Trial: a pilot study. *J Crit Care*. 2005; 20(4): 364–372.
11. Bacchetti P. Current sample size conventions: flaws, harms, and alternatives. *BMC Med*. 2010; 8: 17.
12. van Belle G. *Statistical Rules of Thumb*. New York: John Wiley and Sons, 2002.
13. Julius SA. Sample size of 12 per group rule of thumb for a pilot study. *Pharm Stat*. 2005; 4: 287–291.
14. Carter RE. Application of stochastic processes to participant recruitment in clinical trials. *Control Clin Trials*. 2004; 25(5): 429–436.
15. Carter RE, Woolson RF. Statistical design considerations for pilot studies transitioning therapies from the bench to the bedside. *J Transl Med*. 2004; 2(1): 37.
16. Carter RE, Woolson RF. Safety assessment in pilot studies when zero events are observed. In Auger J-L, ed. *Advances in Statistical Methods for the Health Sciences: Applications to Cancer and AIDS Studies, Genome Sequence Analysis, and Survival Analysis*. Boston: Birkhäuser, 2007; xl: 540.
17. Browne R. On the use of a pilot sample for sample size determination. *Stat Med*. 1995; 14: 1933–1940.
18. Wells G, Beaton D, Shea B, et al. Minimal clinically important differences: review of methods. *J Rheumatol*. 2001; 28(2): 406–412.
19. Chuang-Stein C, Kirby S, Hirsch I, Atkinson G. The role of the minimum clinically important difference and its impact on designing a trial. *Pharm Stat*. 2011; 10(3): 250–256.
20. Kirby S, Chuang-Stein C, Morris M. Determining a minimum clinically important difference between treatments for a patient-reported outcome. *J Biopharm Stat*. 2010; 20(5): 1043–1054.
21. Copay AG, Subach BR, Glassman SD, Polly DW, Jr., Schuler TC. Understanding the minimum clinically important difference: a review of concepts and methods. *Spine J*. 2007; 7(5): 541–546.
22. 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.
23. Geller NL. Design of phase I and II clinical trials in cancer: a statistician's view. *Cancer Invest*. 1984; 2(6): 483–491.
24. Estey EH, Thall PF. New designs for phase 2 clinical trials. *Blood*. 2003; 102(2): 442–448.
25. Lancaster GA, Dodd S, Williamson PR. Design and analysis of pilot studies: recommendations for good practice. *J Eval Clin Pract*. 2004; 10(2): 307–312.
26. Song F, Parekh S, Hooper L, et al. Dissemination and publication of research findings: an updated review of related biases. *Health Technol Assess*. 2010; 14(8): iii, ix-xi: 1–193.