



Pilot Studies What they are and what they are not?

Sandra Taylor, Ph.D.
October 12 & 19, 2016

UCDAVIS
CLINICAL and TRANSLATIONAL
SCIENCE CENTER

UCDAVIS
MIND INSTITUTE

UCDAVIS
COMPREHENSIVE
CANCER CENTER

UCDAVIS
ENVIRONMENTAL HEALTH
SCIENCES CENTER

We are video recording this seminar so please hold questions until the end.

Thanks





I run the biostatistics office hours over at the CTSC where folks can drop in and get quick advice on anything biostatistics related. Commonly, investigators will come in seeking advice on how to analyze their data. After getting the basics about the study design, outcomes, predictors and hypotheses the investigator wishes to evaluate, we typically ask how large a sample there was. Sometimes the investigator rather sheepishly says something like we only had 8 subjects, 4 in each group and then caveat/defend that number saying that it was just a pilot study.

In situations like this, I am not privy to the backstory of the basis and impetus for doing the study so possibly my gross generalization is in error but my general sense is that investigators commonly consider studies with small sample sizes to be pilot studies. Or perhaps more specifically, studies with smaller sample sizes than they think they should have are characterized as pilot studies. This characterization seems to be used to justify the small sample size. As we shall see, while pilot studies do typically have small sample sizes, whether a study is pilot study or not, depends not on the sample size but rather on the objectives of the study.

Objectives

- **Define what a pilot study is**
 - **Highlight uses of pilot studies**
 - **Distinguish between pilot studies and small sample size studies**
- 

What is a pilot study?

- 1. test of the methods and procedures to be used on a larger scale if the pilot study demonstrates that the methods and procedures can work**
- 2. investigation designed to test the feasibility of methods and procedures for later use on a large scale or to search for possible effects and associations that may be worth following in a subsequent larger study**
- 3. Pilot studies are a smaller version of the main study used to test whether the components of the main study can all work together. It is focused on the processes of the main study...**

The first place to start in understanding what a pilot study is, is with some definitions. Here I have three from three different sources. You can find others.

When we look at these definitions, what are common elements of each definition?

What is a pilot study?

1. **test of the methods and procedures to be used on a larger scale if the pilot study demonstrates that the methods and procedures can work**
2. **investigation designed to test the feasibility of methods and procedures for later use on a large scale or to search for possible effects and associations that may be worth following in a subsequent larger study**
3. **Pilot studies are a smaller version of the main study used to test whether the components of the main study can all work together. It is focused on the processes of the main study...**

Each of these definitions, focus on testing methods and procedures.

Note that nowhere in these definitions do we see “testing hypotheses” or “estimating effect size”. The closest we come is in Definition 2 which allows searching for possible effects and associations.

That the focus of a pilot study is on methods and procedures rather than hypothesis testing is critical to recognize and understand.

Focus of pilot studies

- **Assessment of the feasibility of methods**
- **Support refinement of methods prior to implementation of larger study**
- **Not for testing statistical significance *unless it was powered to do so***

In short, pilot studies are used to assess feasibility of a proposed study or study elements and to provide information to support refinement of the study design

Pilot studies are NOT for hypothesis testing (i.e., significance testing) unless the study was powered to do so. Which brings me back to my initial introduction in which I commented that sometimes investigators come in with a small sample size study, characterize it as a pilot and ask use which statistical tests to do. If they wanted to do hypothesis testing they should have done sample size calculations up front. If it is truly a pilot study then it probably isn't appropriate to be doing any statistical testing.

Small studies \neq Pilot study

- **Small study is doable, maybe a student project**
- **Get some data so analyze it**
- **Something is statistically significant**
- **Desire to publish**
- **Investigator knows it is small and not definitive so calls it a pilot study**

Here's what I think happens sometimes. Small studies are more doable in terms of time, money and other resources. Perhaps there is a student who wants a project. So their advisor suggests doing a small study and they get some data. Once there is data there is a desire to analyze it. So they do. They run some statistical tests, look at some plots and lo and behold they find something interesting. Maybe it was what they were hoping for, maybe it was something else. At any rate, now they have an interesting finding and of course there is a desire to publish. Problem is that they know the sample size is small. The "solution" then is to call it a pilot study and that makes it ok.

Problems with this approach

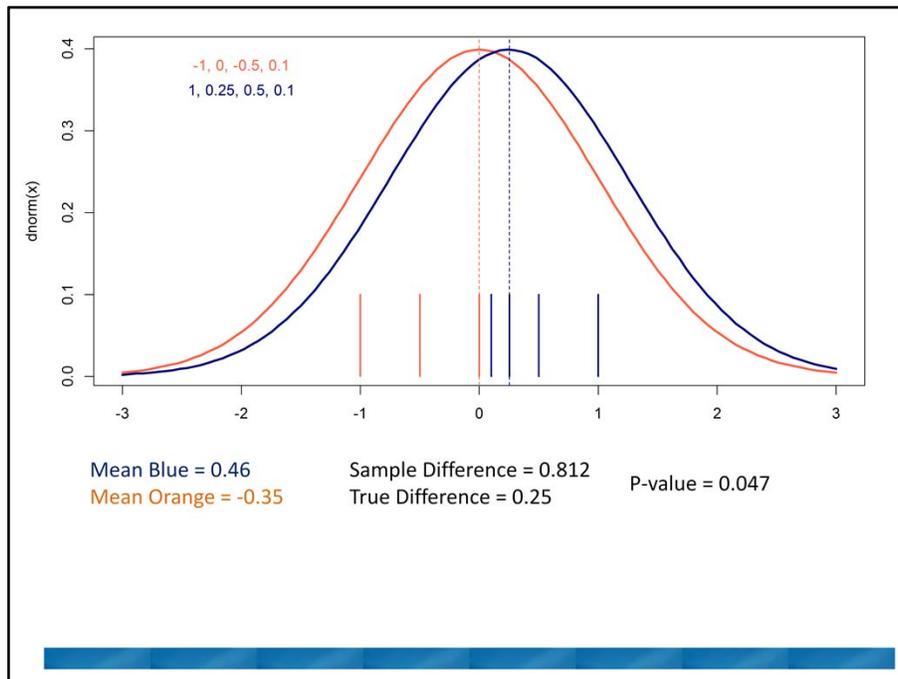
- **You might have gotten “lucky”**
 - 5% of the time when there is no difference you will find significance. If you conducted multiple tests, even more likely that you would find something “significant”
- **Potential to overestimate effect size**
- **Potential to miss interesting effects**

There are a few problems with this type of approach. First, you might have gotten “lucky”. Given the way we do statistical testing and how we make decisions, 5% of the time you will find a significant result even when there is actually no difference. For more information on this concept, I refer you to the first seminar in this series “In Search of Significance” If you test lots of different outcomes, you are even more likely to find something “significant”.

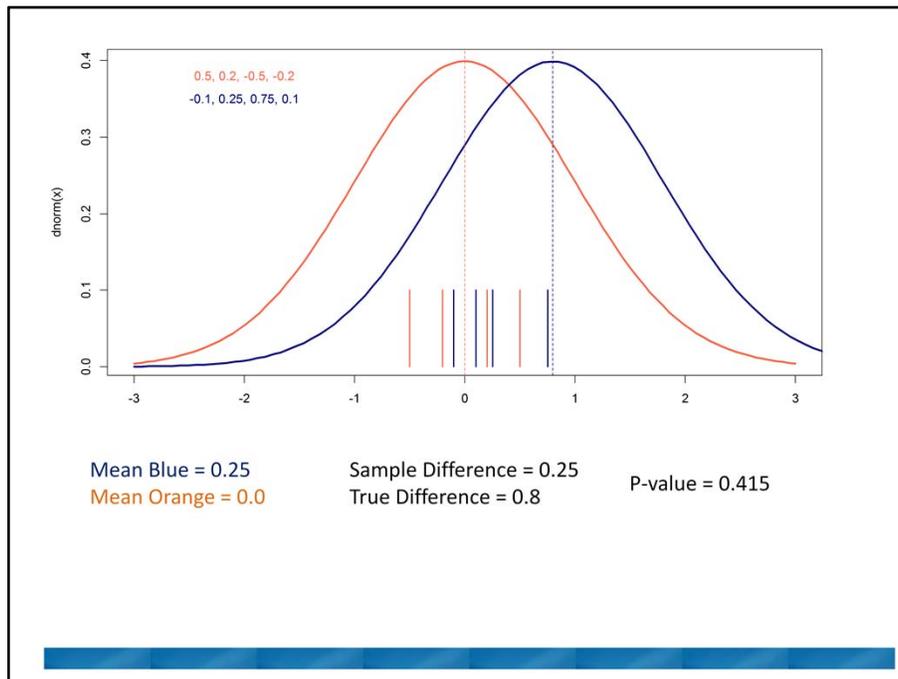
Second, you have the potential to overestimate the effect size.

Third, you also have the potential to miss interesting effects.

Allow me to illustrate these a bit.



Suppose we did a study with 4 subjects in two groups. And suppose that the true difference in means is 0.25. This is illustrated by the orange and blue curves. Now suppose that our samples are as indicated. These are perfectly reasonable values to obtain. None of the values are too wild. If we use a t-test to compare the samples, we find a significant difference in means! We then estimate the treatment effect and we estimate it to be 0.812 from our sample even though the true difference is only 0.25. These are the kind of results that get an investigator excited cause we don't know that the true difference is only 0.25. We think it is 0.812.



Now flip this around. Again suppose we did a study with 4 subjects in two groups. And suppose that the true difference in means is now 0.8. Again the values of the samples we obtained are perfectly reasonable. None of the values are too wild. If we use a t-test to compare the samples, we find no significant difference in means! Our estimated treatment effect is only 0.25. As a result, we decide that the intervention isn't effective and don't pursue it any further.

Problems with this approach

- **You might have gotten "luck"**
 - 5% of the time when there is no effect, you will find significance. If you had used multiple tests, even more likely that you would find something "significant"
- **Potential to overestimate effect size**
- **Potential to miss interesting effects**

NOT A PILOT STUDY

The overarching issue in this scenario is that the study was not designed and powered for significance testing and calling it a pilot doesn't make it ok.

Why conduct a pilot study?

- **Process – assess the feasibility of the steps that need to take place as part of the main study**
- **Resources – assess time and budget requirements**
- **Management – potential human and data optimization problems**
- **Scientific – Assess safety, determine of dose levels and variance of parameters**

Let's back up. Recall the definitions from earlier. The focus of a pilot study are processes and methods. So what can we get from pilot studies?

Process – we can assess feasibility. In other words, the pilot study can be a dry run where you can figure out what works, what doesn't work, how to improve the study process for smooth and efficient data collection

Resources – assess time and budget requirements. Really good to have when you are preparing a full blown study

Management – people and data management. Work out the kinks.

Scientific – can be used to assess safety, determine dose levels, estimate variance. Some caution here in that you need to power it to do this

Processes

- **Determine rates of recruitment, retention, refusal, failure/success rate, compliance/adherence rates**
- **Assess/refine eligibility criteria**
 - Is it clear who is and is not eligible?
 - Are criteria too restrictive or liberal?
- **Evaluate study questions/data collection tools**

Let's drill into each of these areas in more detail to illustrate the types of questions that pilot studies can and should be used to address

Rates of recruitment is a really critical piece of information and often underestimated, sometimes by a lot.

Resources

- **Time to collect data/complete surveys**
- **Equipment accessibility/availability**
- **Software capability**
- **Site capacity and ability to implement study**

Underestimating these items can result in significance overruns.

Management

- **Identify center and staff challenges with managing the study**
- **Data collection and storage capacity**
- **Correct and adequate data collected**

Piloting to assess management issues is particularly important if you are running a multi-center study. Multi-center could refer to different geographic locations or different units at the same facility, e.g., different ICUs at UC Davis. Each may run things a little differently and you want to make sure that the processes will work adequately in all units. Also, you might need to make sure units/centers can't access data from other units/centers

Scientific

- **Determine if safe to use drug/intervention**
- **Determine safe dose level**
- **Determine if subjects respond**
- **Estimate treatment effect**
- **Estimate variance of treatment effect**

There may be circumstances under which you want a pilot study to evaluate various scientific questions. This is where I think things get a little fuzzy between a pilot study and let's call it a hypothesis testing study and sample sizes need to be considered carefully here. Recall my previous example of effect size estimation based on a very small sample size. We will revisit this in a bit.

So, now that we have a good understanding of what pilot studies should be used for, let's be clear about what they can't or shouldn't do.

What pilot studies can't do

- **Hypothesis testing**
 - It should not be considered a preliminary test of the intervention hypothesis. No hypothesis testing because "the limited state of knowledge about the methods or intervention in the patient population to be studied and the smaller proposed sample size."
- **Tolerability and Safety**
 - Only if someone dies would there be some info on this from a pilot study due to the small sample size. Nevertheless if some adverse event concerns do arise in a pilot study the event rates by groups with 95% CIs should be reported.

If want to use pilot study to assess tolerability and safety, you need to have an appropriate sample size.

When do I need a pilot study?

- **What do I need to know to be able to successfully execute a definitive study?**
- **What parts of the design do I say, “Gosh, I just don’t know”?**
- **What did I wish I knew before starting my last study?**
- **What are the major uncertainties that could screw this whole thing up?**

With this new/improved understanding of what a pilot study, you might be questioning if you would ever need to do one as well as thinking you would rather put your resources into the primary study than spend time and money doing a pilot. Fair enough. You certainly don't need to pilot every study.

Here are some questions that you might ponder when deciding whether something needs piloting.

What do I need to know to be able to successfully execute a definitive study? If you are proposing rat study and you have done many similar rat studies then you probably have a good handle on what it takes to be successful. Alternatively if you are implementing a multi-center clinical trial and you have never done this before, piloting would be very informative.

If there is a part of your study where you think “I really don’t know how this is going to play out” – that’s a good thing to pilot.

You should really take a hard look at and challenge your assumptions. Discuss the project with colleagues. We tend to be optimistic and we want to paint a rosy picture for funding agencies.

Many pilot studies not well designed

- **No clear feasibility objectives**
- **No clear analytic plans**
- **No clear criteria for success of feasibility**

This is the “Let’s just see what happens” and “We’ll know what to do when we see the results” approach.

Perhaps after hearing this talk you decide that you need to pilot some aspects of a study you want to conduct. The next obstacle to gaining the full value and benefit of pilot studies results from the studies being poorly designed. Again I think there is a misperception about pilot studies and they are not taken seriously or regarded as needing to be as rigorous as a “real” study. For the results of ANY study to be useful it must be designed properly.

You limit the value of the pilot study if you don’t clearly define you objectives, the analytic plan and the decision criteria.

Take pilot studies seriously

- **Design it to get the information you need**
- **Define your objectives**
- **Specify an analytical plan**
- **Define criteria and how you will use results**

So, if you decide to pursue a pilot study, DO take it seriously. After all you are going to spend time, money and resources on the pilot study so take the time to design and implement it properly.

What does that mean?

1. Identify the information you need (be specific) and make sure the study is designed to give you this information
2. Define your objectives. What specifically do you need at the end of the study?
3. Prepare an analytical plan. How are you going to summarize/analyze the information you get? This doesn't have to be complicated and shouldn't be an extensive statistical analysis. Still think about this.
4. Define criteria and how you will use the results – This goes along the lines of little thought experiments. If I see X, what will that tell me. What decision will I make based on that? What if I see Y? and so on.

Sometimes folks come into office hours and say that they are planning a pilot study. Usually they come in asking about sample size and they seem rather lost when we start asking them questions regarding what decisions they are going to make depending on the results

Outcome Decisions

- **Stop – main study not feasible**
- **Continue, but modify protocol – feasible with modifications**
- **Continue without modifications but monitor closely – feasible with close monitoring**
- **Continue without modifications – feasible as is**

You have 4 possible choices of what to do following a pilot study.

1. Stop
2. Continue with modifications
3. Continue without modifications but monitor
4. Continue without modification

Recognizing what the decision options are can help you in developing an analytical plan and setting criteria.

Sample Size Justification

- **No hypothesis testing sample size calculation often unneeded**
- **Still, need a sample size *justification***
 - Pragmatics of recruitment
 - Necessities for examining feasibility
- **Desired precision of estimates or preliminary effect size estimates**
 - Can be addressed with sample size calculations

The last topic I want to touch on with respect to pilot studies is Sample Size Justification. Sometimes folks come to us for help with determining sample size for a pilot study. If we think about pilot studies as defined here, they are focused on processes and methods, and not on hypothesis testing or estimation. In these cases, there is no sample size *calculation*. Sample size calculations are determined by wanting to be able to detect a certain magnitude of difference or estimate a quantity with a certain precision. Nevertheless, you will need a sample size *justification*. The justification may be based on the pragmatics of recruitment, what is necessary for examining feasibility. These are not statistical questions.

Now if you need/want to estimate some parameters (e.g., effect sizes, standard deviation) with a certain precision in order to plan a larger study, that can and should be determined with sample size calculations.

Pilot Data for Sample Size Estimation

- **Pilot data often used to estimate sample size for subsequent study**
- **Potentially problematic**
 - Imprecise estimates due small sample size
 - Overly optimistic effect size estimates
 - “Underly” optimistic effect size estimates

There is a long tradition of using estimates from pilot studies, e.g., standard deviations, effect sizes, to determine sample size requirements for the subsequent definitive study. While I don't want to discourage collecting and making use of pilot data, I do want to point out some concerns. First, because of the typically small sample size, parameter estimates from pilot studies, may not be very precise. Recall the estimates of effect size we had with the example of 4 subjects per group.

Second, if you are moving forward with a study because of “promising” results in the pilot study, the effect sizes may be overestimated as we saw in the prior example. This could lead to an underpowered study to detect the real difference leading to a “negative trial outcome”.

Alternatively, you could have a small estimated effect size (a false negative) from a pilot study and either a) not pursue a larger trial or b) “overpower” the subsequent trial leading to extra expense, time, and exposure of subjects to the trial conditions.

Impact of Imprecision

- **Consider 8 subjects/group**
- **Assume standard deviation of 1**
- **Suppose estimate difference in mean as 1 (i.e., 1 SD difference)**
- **95% CI for effect, e.g. [0.4, 1.6]**
 - If 0.4, need 132/group for 90% power
 - If 1.0, need 22/group for 90% power
 - If 1.6, need 9/group for 90% power

Sample size calculations can be strongly affected by the assumed effect size.

Recommendations

- **If pilot data is intended to yield estimates for sample size calculations, power it adequately**
- **Consult published works for effect sizes and standard deviations**
- **Use clinically meaningful effect size**
- **Err on the side of caution**

Here are some recommendation with respect to using pilot data for sample size calculations. First, if you are doing a pilot study with the express objective of getting estimates to power a larger study, then be sure your pilot study is large enough to give adequate precision. Second, in all cases, consult published works. If it is a common outcome and you are interested in whether your new intervention is better than standard of care, then you probably want it to have an effect at least as big as the standard of care and this can be used as effect size. Published works can help provide some context and general magnitude of the effects you might see. Third, regardless of what the pilot data show, always think about what is a clinically meaningful difference. You can think of this in terms of how big of a difference do I need to see before I would change the way I am doing something? Lastly, you can err on the side of caution. From the pilot data you might estimate an effect size of 1 like in our example but you might want to power the larger study to detect a difference as small as 0.5 given the uncertainty of the effect size estimate.

Summary Points

- **Small study \neq Pilot study**
 - **Pilot studies focus on assessing and refining methods and processes *before* undertaking a large definitive study**
 - **Take pilot studies seriously and design them to achieve specific objectives**
- 

Help is Available

- **CTSC Biostatistics Office Hours**
 - Every Tuesday from 12 – 1:30 in Sacramento
 - Sign-up through the CTSC Biostatistics Website
- **EHS Biostatistics Office Hours**
 - Every Monday from 2-4 in Davis
- **Request Biostatistics Consultations**
 - CTSC - www.ucdmc.ucdavis.edu/ctsc/
 - MIND IDDRC - www.ucdmc.ucdavis.edu/mindinstitute/centers/iddrc/cores/bbrd.html
 - Cancer Center and EHS Center

References

- Leon et al. 2011. The role and interpretation of pilot studies in clinical research. *J Psychiatr Res.* 45(5):626-629.
- Thabane et al. 2010. A tutorial on pilot studies: the what, why and how. *BMC Medical Research Methodology* 10:1.
- Hertzog, M.A. 2008. Considerations in determining sample size for pilot studies. *Research in Nursing & Health* 31:180-191
- Julious, Steven A. 2005. Sample size of 12 per group rule of thumb for a pilot study. *Pharmaceutical Statistics* 4: 287-291.