Statistical Review and Design of Pilot Studies

Hae Kyung Im
Kristen Wroblewski

Outline

- Definition of pilot studies
- Reasons to conduct a pilot study
- Reviewer's checklist
- Example
- On using pilot data for sample size calculation
- Internal vs. external pilots

Pilot Study Abuses

- Term is overused
- Studies that look at feasibility and those that do hypothesis-testing on a small group of subjects are lumped together and labeled pilot studies
- Pilot and Collaborative Translational and Clinical Studies funding as part of the CTSA

Small Exploratory Study

- These types tend to be labeled as pilot studies and researchers claim there is no need for sample size justification
- The goals go beyond testing feasibility
  - Hypothesis generating study
    - No need to assess power
    - A few subjects should suffice (even n=1)
  - Hypothesis testing study
    - Only large effect sizes will be detected
    - Analysis of power and sample size is needed

(What is a pilot study, Paul Stewart)

Pilot Study

- A small investigation to test the feasibility of procedures and to gather information prior to a larger study
  - Designed to assess whether study is worth pursuing and what changes need to be made

Reasons for conducting pilot studies

- Study administration
- Resources and data management
- Scientific

(Thabane et al. 2010 BMC Medical Research Methodology)

Study Administration Reasons

- Recruitment, refusal, and retention rates
- Compliance/adherence rates
- Are eligibility criteria clear and reasonable?
- Are study questionnaires well understood?
- Are instructions to investigators clear?

(Thabane et al. 2010 BMC Medical Research Methodology)

Resource and Data Management Reasons

- Are all study sites willing and capable of carrying out the study?
- Are there sufficient resources dedicated to patient and data management?
- Can data from different sources be merged successfully?
- Are all important variables being collected?

(Thabane et al. 2010 BMC Medical Research Methodology)

Scientific Reasons

- How much variability is there in the outcome(s)? What is the effect size?
- Choosing between multiple outcome measures based on reliability/feasibility
- Are assay methods appropriate?

(Thabane et al. 2010 BMC Medical Research Methodology)

A pilot study is NOT…

An excuse to use a small number of patients and not consult with a statistician due to limited funds or time

Checklist for reviewers (and investigators) Part 1

First, determine whether it's truly a pilot study:

- At least one of the reasons (study administration, data management, scientific) should apply.
- No hypothesis should be tested.

(What is a pilot study, Paul Stewart)

Checklist for reviewers (and investigators) Part 2

- Are aims and objectives clearly stated?
- Is the sample size justified?
- Have they addressed how the data collected will be used in the design of a larger study?
- Will this study answer the question of whether a full scale trial/experiment is worth pursuing?
- Are there clear criteria that will lead to the decision of pursuing a larger study?

(What is a pilot study, Paul Stewart)

Example

- Prior to a larger clinical trial, a pilot study involving 10 patients is proposed. The goal of the project is to investigate how well patients will tolerate wearing a new ambulatory heart monitor while receiving an experimental medication. Data will be downloaded from the monitors and will be analyzed using a t-test procedure for comparison of post-treatment heart rate to pre-treatment heart rate.

(What is a pilot study, Paul Stewart)

Example

- Main goal was to assess tolerability but there is no measure defined for it. No data to address the main goal is collected.
- There is an implicit hypothesis tested.
- Sample size is not justified and may be driven by the desire to detect a difference.

(What is a pilot study, Paul Stewart)

Use of pilot study for sample size determination

- Sample variance from a pilot study is positively skewed
- So more than 50% of the time the sample variance will be lower than the true variance
- How often does the calculated power reach or exceed the actual power?
  - Using sample variance
  - Using 100(1-γ) percent upper one-sided confidence limits

(Browne R, 1995 Statistics in Medicine)

Let $X_i \quad i = 1, 2, ..., n$ $\qquad X_i \sim N(\mu, \sigma^2)$

$H_o : \mu = \mu_o \qquad H_1 : \mu = \mu_1 > \mu_0$

$t(\delta, \nu) = \sqrt{n}(\bar{x} - \mu_o)/\sigma$ $\qquad$ noncentral $t_{\delta,\nu}$

$\delta = \sqrt{n}(\mu_1 - \mu_o)/\sigma$

$P(\ t(\delta, \nu) \geq t_{1-\alpha,\nu}\ ) \ \geq \ 1 - \beta$

σ is unknown

$S_m =$ sample variance or $S_m \frac{m-1}{\chi^2_{m-1,1-\gamma}} = 1 - \gamma$ UCL

How often do we achieve the planned power?

Table I. Relative frequency with which actual power equalled or exceeded planned power of 0·80 in 2000 trials, when using indicated estimator of σ in determining sample size for a two-tail one-sample t-test

Effect size*	Pilot sample size	Estimator of σ used in sample size determination					
		Pilot sample standard deviation	50% UCL† on σ	60% UCL† on σ	70% UCL† on σ	80% UCL† on σ	90% UCL† on σ
0·10	5	0·39	0·50	0·60	0·70	0·80	0·90
0·10	10	0·43	0·50	0·60	0·71	0·80	0·91
0·10	30	0·48	0·51	0·61	0·71	0·82	0·91
0·10	50	0·48	0·51	0·62	0·71	0·81	0·90
0·10	100	0·50	0·52	0·61	0·71	0·82	0·92
0·40	5	0·43	0·51	0·61	0·70	0·79	0·89
0·40	10	0·48	0·54	0·65	0·76	0·84	0·93
0·40	30	0·54	0·57	0·66	0·76	0·83	0·92
0·40	50	0·56	0·58	0·67	0·75	0·84	0·92
0·40	100	0·61	0·56	0·72	0·81	0·89	0·95
0·75	5	0·52	0·60	0·69	0·78	0·87	0·94
0·75	10	0·62	0·66	0·74	0·83	0·88	0·95
0·75	30	0·75	0·75	0·83	0·88	0·93	0·98
0·75	50	0·83	0·82	0·87	0·93	0·94	0·99
0·75	100	0·91	0·90	0·92	0·95	0·95	1·00

Types of Pilot Studies

- External
  - Distinct from larger study
  - Test of feasibility
  - Small sample
- Internal
  - Two-stage design with first stage deemed pilot phase
  - Permits refinement of parameters used in initial study design by re-estimating at end of first stage
  - Larger sample size with little increase in cost/time

Sample size for internal pilots

- Variety of methods (Stein 1945, Wittes and Brittain 1990, Gould and Shih 1992)
- These methods behave similarly when the interim sample size is large (greater than 40)
- Uncertainty on how to choose fraction of patients used in first, pilot phase

(Zucker et al. 1999 Statistics in Medicine)

Wittes-Brittain Method

- Estimate the sample size per group, n, as usual using preliminary estimate of variance, τ^2.
- Select proportion, p, so the first pn in each group comprise the pilot phase.
- Estimate variance, s^2, after pilot phase is complete.
- If $s^2 \leq \tau^2$, then continue as planned. Otherwise, adjust the sample size using s^2
- Inflation of true α-level will be small in most cases

(Wittes et al. 1990 Statistics in Medicine)

Summary

- Pilot studies (feasibility) vs. small exploratory studies (hypothesis generation or testing)
- Checklist for statistical review
- Sample variance vs. upper confidence limits
- Internal vs. external

References

- Lancaster et al. (2004) *Design and analysis of pilot studies: recommendations for good practice* Journal of Evaluation in Clinical Practice 10:2, 307-12.
- Thabane et al. (2010) *A tutorial on pilot studies: the what, why and how* BMC Medical Research Methodology 10:1
- Wittes J and E Brittain (1990) *The role of Internal Pilot Studies in Increasing the Efficiency of Clinical Trials* Statistics in Medicine 9, 65-72.

References (continued)

- Zucker et al. (1999) *Internal Pilot Studies II: Comparison of various procedures* Statistics in Medicine 18, 3493-2509.
- Browne RH (1995) *On the use of a pilot sample for sample size determination* Statistics in Medicine 14, 1933-40.
- Stewart PW, "What is a Pilot Study?" http://www.cincinnatichildrens.org/research/cores/gcrc/protocols/Small_or_Pilot_Studies.htm

Questions?