

# Designing *Your* Research Study

Essential concepts, Best practices, Pitfalls, Speedy IRB approval



Paul W. Stewart, PhD

Professor of Research in Biostatistics

# Overview

---

Let's talk about designing *your* research study

- Essential concepts
- Best practices
- Pitfalls to avoid
- Speedy IRB Approval

We will do this in the context of some of the more challenging kinds of studies to design and execute: “Pilot” studies and observational studies.

The ideas presented apply to all kinds of research studies.

- **Challenges in Pilot Studies**
- Challenges in Observational Studies
- Designing Studies
- Choosing a Sample Size
- Summary: Strategies for Speedy IRB Approval
- Appendix (a simulation)

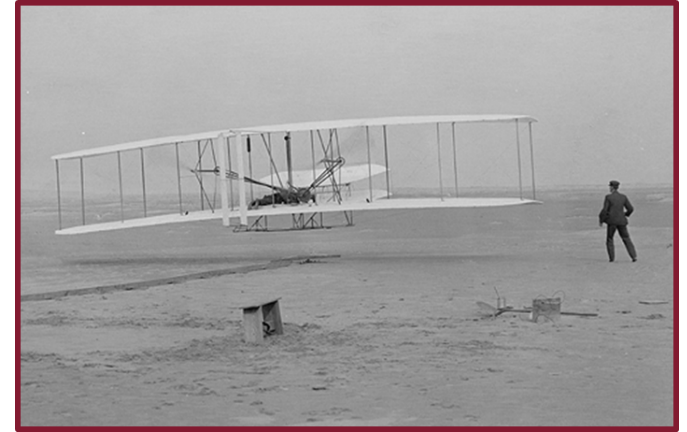
# 1903 Flight: Was it a Pilot Study ?



# 1903 Flight: Was it a Pilot Study ?

## Check List

- ✓ Little funding; no federal grants
- ✓ Winging it on design and analysis
- ✓ Uncertain expectations about the project
- ✓ Not assisted by the “leading experts” of the time
- ✓ Sample size:  $N = 3$  controlled flights on December 17, 1903
- ✓ Did not consult a biostatistician
- ✓ News Editors were not initially interested in the results



# 1903 Flight: It Was a Necessary First Step

The Wright brothers' historic achievement of controlled flight in 1903 was **a necessary step** in a long sequence of steps.



# The Role of Small Preliminary Studies

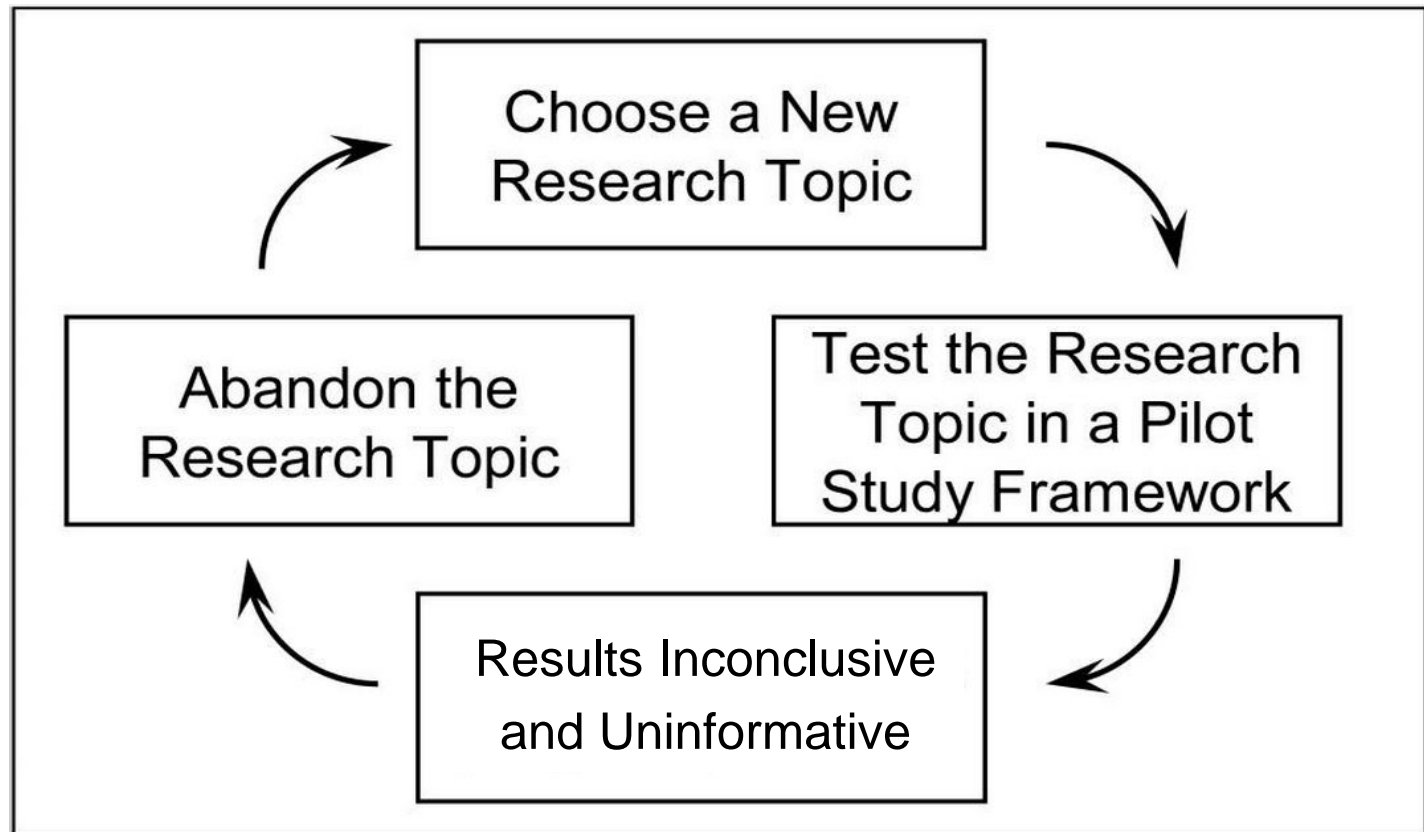
---

## Reasons we do small-scale preliminary studies

- To obtain information needed to plan a new study
  - ◆ to study feasibilities, costs and time requirements.
  - ◆ to generate new hypotheses.
  - ◆ to report preliminary data in a future grant proposal.
- It would be foolish to start the new study without that information
- The preliminary study can be a career-building achievement.
- Other reasons:
  - ◆ need practice in doing clinical studies (it is a training exercise)
  - ◆ seed-grant funds are available ! (even if extremely limited).

# Career-building? Your Personal Research Risk

Poorly designed “pilot studies” can harm your career if they prove to be uninformative, inconclusive, or misleading.





# Pilot Study Challenges

---

## Pilot studies are often disadvantaged

- Lack of funding to support collaboration with specialists, statisticians, data management professions tends to reduce the quality of all aspects of the research.
- That's bad because... Pilot studies are NOT easy to design well.
- Specialists, statisticians and others may not be interested in collaboration if there will be no funding, no manuscript, and no grant proposal.
- Often, sample size (and team size) must be small because of a lack of funding.

# Pilot Study Challenges

---

**The “Pilot Study” label is associated with problems ...**

- ◆ Little or no funding
- ◆ The study being poorly designed
- ◆ Mis-alignment of aims / design / analysis plans
- ◆ Low expectations for the project
- ◆ Lack of careful planning of the protocol details
- ◆ Sample size (N) seems too small or too large
- ◆ Little or no justification given for the choice of N
- ◆ Lack of considerations of statistical strategy & methods
- ◆ Lack of an adequate plan for research data management

# Pilot Studies Deserve Better

---

## Common Misconceptions about Pilot Studies

- ◆ Pilot studies are relatively easy to design well
- ◆ Study is unfunded and badly designed, but that is “not a problem”
- ◆ Plans for ensuring data quality are not necessary
- ◆ A well-considered strategy for analyzing data is not necessary.
- ◆ Statistical methods are not applicable or not necessary.
- ◆ For “obvious” reasons justification of sample size is not needed
- ◆ Study is believed to be under-powered, but that is “not a problem”
- ◆ It is not necessary to know details of the “next” future study

# Preliminary Studies: Meaningful Definitions

---

## **Exploratory Study**

An investigation for generating hypotheses about the target population, based on a sample.

## **Small Risky Inferential Study**

A limited investigation of the scientific research questions about the target population, based on a sample.

## **Preparatory / Pilot-Testing / Feasibility Study**

An investigation of the performance characteristics of “the next protocol” for the target population, based on a sample.

# Preliminary Studies: Meaningful Definitions

---

## Exploratory Study

An investigation for generating hypotheses about the target population, based on a sample.

For example ...

- A case study ( $N = 1$ )
- Data mining ( $N = \text{big}$ )
- Scatter plots to visualize relationships among variables
- Cluster analysis: searching for patterns
- Model building (e.g., lasso regression, random forests)

**Some methods are better than others for generating hypotheses.**

# Preliminary Studies: Meaningful Definitions

---

## Small Risky Inferential Study

A limited investigation of scientific research questions about the target population, based on a sample.

For example ...

- Go-NoGo: does the treatment effect confidence interval include clinically important magnitudes of effect?
- Testing the primary null hypothesis (power may be low)
- Testing 1000 null hypotheses (risk of misleading results)
- Rough estimation of treatment effects (low precision)
- Rough estimation of a correlation coefficient (low precision)

Your personal research risks ...

- The study may be misleading or inconclusive

# Preliminary Studies: Meaningful Definitions

---

## **Preparatory / Pilot-Testing / Feasibility Study**

An investigation of the performance characteristics of “the next protocol” for the target population, based on a sample.

### **Will the future protocol be successful?**

- The focus is on making sure it will be successful.
- Knowledge about the future protocol is required.

### **For example ...**

- Verify that a procedure is reliable and tolerable
- Obtain estimates needed to plan the future study.
- Pilot-test (“debug”) an assay, tool, or data-entry system.
- Evaluate validity & reliability of a questionnaire.

# Preliminary Studies: Meaningful Definitions

---

## Preparatory / Pilot-Testing / Feasibility Study

**For example, evaluate feasibility in terms of ...**

- recruitment rate that can be expected
- incidence rate of refusal
- incidence rate of drop-out
- protocol adherence rate that can be expected
- demonstrated ability to perform procedures
- demonstrated ability to collect data
- costs and time requirements
- validity / reliability / accuracy of new measures

**We use a sample to obtain point- and interval- estimates of these population parameters.**



# Preliminary Studies: Meaningful Definitions

---

Most studies have a mix of aims ...

## **Exploratory Study**

An investigation for generating hypotheses about the target population, based on a sample.

## **Small Risky Inferential Study**

A limited investigation of the scientific research questions about the target population, based on a sample.

## **Preparatory / Pilot-Testing / Feasibility Study**

An investigation of the performance characteristics of “the next protocol” for the target population, based on a sample.

# Most Studies Have a Mix of Aims

---

## Example: Preliminary Study with Mixed Aims

- Aim 1: Small Exploratory Study (initial data for grant proposal)
- Aim 2: Pilot-Testing (find and correct problems in procedures)
- Aim 3: Feasibility Study (evaluate tolerability and retention)
- Aim 4: Feasibility Study (demonstrate ability to perform tasks)

These aims differ in regard to

- Design issues,
- Analysis strategy,
- Data management requirements,
- Sample size considerations.

# Most Studies Have a Mix of Aims

---

## Example: **What Kind of Study is This ?**

- Aim 1. A well-designed RCT to evaluate efficacy & safety.
- Aim 2. Exploratory analyses for biomarkers and outcomes.
- Aim 3. Small risky inferential study of a very small subgroup.
- Aim 4. Pilot-testing and feasibility study of a new assay.

Q: Is my study a “pilot study” ?

A: That may be the wrong question. It is more useful to focus on providing meaningful descriptions of your specific aims.

# Problems with Small Risky Inferential Studies

## Low Precision: Truth Inflation

If the population treatment effect is small but clinically important, and if the sample size (N) is small so that precision is low, then ...

- ◆ the statistical estimate of treatment effect will be highly variable
  - ◆ the confidence intervals will be very wide
  - ◆  $H_0$  “*the treatment effect is exactly zero in the target population*”
  - ◆ P-value  $< \alpha = 0.05$  when a sample of patients is drawn such that the treatment effect estimate happens to be huge
- 
- This is known as *truth inflation* or the *winner's curse*.
  - Reporting only the huge effect and p-value will be misleading.
  - This is a problem in fields where many researchers compete to publish the most exciting results and the focus is on p-values.

# Problems with Small Risky Inferential Studies

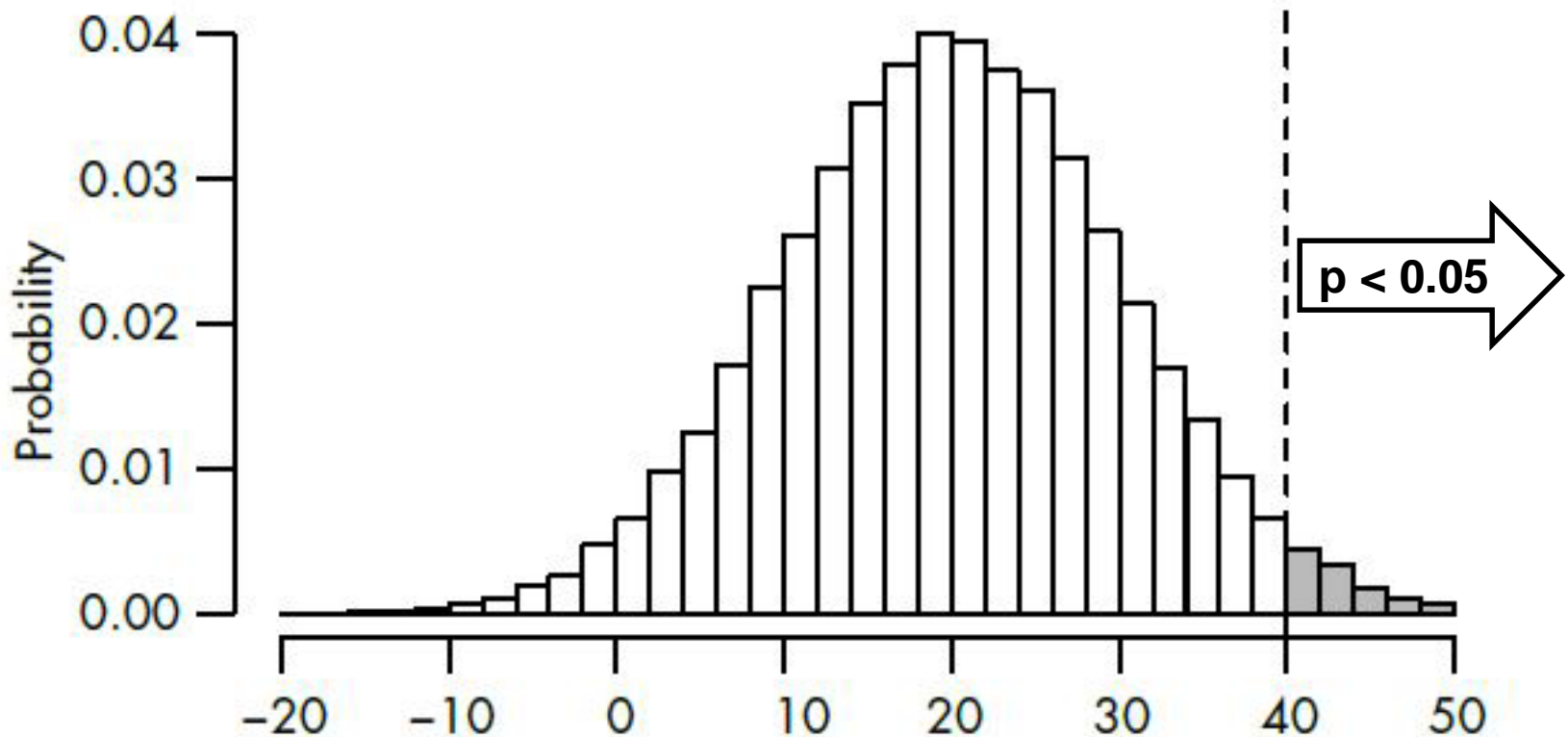
---

## The cure for “Truth Inflation”

- The results will be misleading only when there is over-reliance on the p-values and a failure to appropriately focus instead on the point- and interval- estimate.
- A look at the 95% confidence interval would clarify that the estimator has very low precision. The lower limit of the very wide interval will be close to zero –but the interval will not actually include zero.
- The C.I. shows that it is entirely plausible that the true magnitude of the effect may be very small or very large.
- $P\text{-value} < \alpha$  (correctly) establishes that the effect is not zero

# Problems with Small Risky Inferential Studies

Truth Inflation:  $p < 0.05$  only when the estimate is extreme



**Estimate of Treatment Effect**

(depends on the sample of patients enrolled)

--from *Statistics Done Wrong*

# Problems with Small Risky Inferential Studies

Reference for previous figure ...

## Statistics Done Wrong: The Woefully Complete Guide

by Alex Reinhart

March 2015, 176 pp.

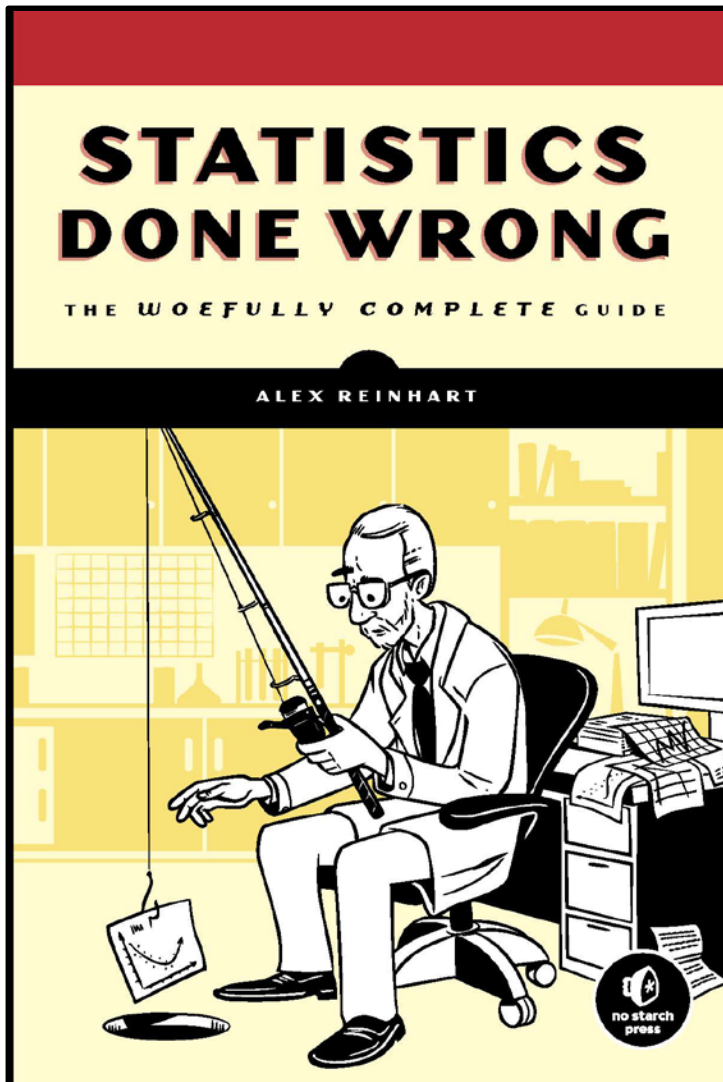
ISBN: 978-1-59327-620-1

\$24.95 Print Book and FREE e-book

\$19.95 e-book (PDF, Mobi, and ePub)

[www.nostarch.com/statsdonewrong](http://www.nostarch.com/statsdonewrong)

Fun to read.



# Problems with Small Risky Inferential Studies

## Low Precision: Unreliable Input for Sample Size Planning

To plan a future RCT of systolic blood pressure (SBP), information is needed about SBP variability among patients in the target population. The population standard deviation ( $\sigma$ ) will be estimated by studying a small sample of  $N_1$  patients. If  $N_1$  is small so that precision is low, then ...

- ◆ the statistical estimate of  $\sigma$  (call it “SD”) will be highly variable
- ◆ the confidence interval for  $\sigma$  will be very wide

Based on  $N_1 = 10$  subjects, suppose  $SD = 14.2$  is the estimate, and we use that number in a power calculation which suggests we need  $N_2 = 24$  subjects in the RCT.

What are your concerns about this scenario ?



# Problems with Small Risky Inferential Studies

## Low Precision: Unreliable Input for Sample Size Planning

$$N_1 = 10 \rightarrow SD = 14.2 \rightarrow N_2 = 24 \text{ for new RCT}$$

### Concerns...

- Small external pilot studies can suggest  $N_2$  values that are extremely far from optimal --either much too large or much too small.
- Even if  $N_1$  is large there is a substantial risk of choosing  $N_2$  badly.
- In practice, large  $N_2$  values are deemed infeasible, but small values of  $N_2$  are deemed easily feasible, what will happen?
  - ◆ The future study will go forward only if  $N_2$  is small.
  - ◆ This selection bias results in an excess of inconclusive studies.

# Problems with Small Risky Inferential Studies

## Low Precision: Unreliable Input for Sample Size Planning

$$N_1 = 10 \rightarrow SD = 14.2 \rightarrow N_2 = 24 \text{ for new RCT}$$

### To avoid that problem...

- Make use of information (e.g., about SD) in previously published studies
  - ◆ Numerous studies of SBP are highly informative even if they were not studies of the novel treatment regimen and subpopulation of interest to you.
- Consider use of
  - ◆ internal-pilot study designs,
  - ◆ group-sequential study designs,
  - ◆ other kinds of adaptive study designs.
- Give serious attention to (perhaps large) uncertainty indicated by the C.I.s for inputs (e.g., SD) and C.I.s for estimates of power and the margin of error. When those C.I.s are very wide, understand that the sample size analysis is highly uncertain.

# Designing 'Pilot' Studies

---

Moore CG, Carter RE, Nietert PJ, Stewart PW (2011).

**“Recommendations for Planning Pilot Studies in  
Clinical and Translational Research.”**

*Clinical and Translational Science*, 4(5):332-337.

# Essential concepts, Best practices, Pitfalls, Speedy IRB approval

---

- Challenges in Pilot Studies
- **Challenges in Observational Studies**
- Designing Studies
- Choosing a Sample Size
- Summary: Strategies for Speedy IRB Approval
- Appendix

# Challenges in Observational Studies

---

Lederer DJ, et al. (August 2018)

“Control of Confounding and Reporting of Results in Causal Inference Studies: Guidance for Authors from Editors of Respiratory, Sleep, and Critical Care Journals.”

*Annals of the American Thoracic Society* vv. pp-pp.

(doi: 10.1513/AnnalsATS.201808-564PS.)

[Epub ahead of print]

# Challenges in Observational Studies

---

## Key principles explicated by Lederer, et al. (2018)

1. Causal inference requires careful consideration of confounding
2. Interpretation of results should not rely on the magnitude of p-values
3. Results should be presented in a granular and transparent fashion; with reference to the STROBE statement and checklist.

-----

STROBE (2007): Strengthening the Reporting of Observational Studies in Epidemiology

# Essential concepts, Best practices, Pitfalls, Speedy IRB approval

---

- Challenges in Pilot Studies
- Challenges in Observational Studies
- **Designing Studies**
- Choosing a Sample Size
- Summary: Strategies for Speedy IRB Approval
- Appendix

# Designing Studies: Best Practices

## Specific Aims

- List all the specific aims.
- Check for a 1-to-1 match between aims and analyses
- Provide complete details for each aim.
- Be creative and thoughtful in describing your aims.  
(Thus, avoid broad ambiguous over-use of “pilot study”)
- Distinguish among different kinds of aims ...
  - ◆ Simple description of the sample.
  - ◆ Use sample to test hypotheses about the target population
  - ◆ Obtain point- and interval-estimates of population parameters
  - ◆ Exploratory: generate hypotheses about the population
  - ◆ Preparatory / pilot-testing / feasibility investigations



# Designing Studies: Best Practices

---

## Justify Design Features

- Provide details and a rationale for the...
  - Treatment design,
  - Observational / Experimental design,
  - Measurement design
- If the experimental design is a crossover, carefully justify the length of the washout intervals
- Provide plans for
  - Randomization and concealment,
  - Stratification and/or Matching,
  - Blinding

# Designing Studies: Best Practices

---

## Database Management Plan

- Provide complete plans for data management that ensure
  - data quality
  - data security
  - data confidentiality
- Collaborate with experts on data management plans

# Designing Studies: Best Practices

---

## Statistical Analysis Strategy and Methods

- Carefully align aims / design / analysis plans
- Clearly define all variables used in the analyses; specify their units of measure or their range of scale
- For each aim, provide a complete data analysis plan
- Confidence intervals should play a major role.
- Collaborate with your friendly local biostatistician

# Designing Studies: Example of Mis-alignment

---

**A “feasibility study” is proposed with N = 10.**

- The only stated aim is to investigate how well patients tolerate wearing an ambulatory heart monitor before receiving and while receiving a new experimental drug.
- No justification was given for the choice of N=10.
- Database plans: download heart monitor data.
- Analysis plans: apply a t-test procedure to compare pre-treatment to post-treatment heart rate variability.
- Stated purpose: study will “determine” the efficacy.

**What is wrong with this picture ? (Statistical concerns)**

# Designing Studies: Example of Mis-alignment

---

## Design & Analysis not aligned with Aims

- No “tolerability” data are to be collected.
- No plans to analyze “tolerability” data.
- No discussion about “tolerability” requirements.
- No plan for drawing a conclusion on “feasibility”.
- Discussion and justification of the choice of N in terms of an appropriate design & analysis plan for studying “feasibility”.
- Part of an analysis plan for a future study was inserted.
- Confusion of the aims of this study with the aims of the entire line of research.
- Overuse of the catch-all word ‘determine’.  
(Better: ‘Estimate’ the treatment effect, or ‘Evaluate’ efficacy.)

# Designing Studies: Example of Mis-alignment

---

## Design & Analysis not Aligned with Aims

In regard to the proposed t-test;

If a “small risky inferential study” aim is intended, then...

- that aim should be stated and appropriate plans and discussion given in the protocol document,
- any intent to publish should be explicitly stated.

# Designing Studies: Example of Mis-alignment

## Confusion of Aims in Research Proposals

- The aims of the entire line of research
  - ◆ Develop a safe and effective treatment for disease “X”
  - ◆ Motivation: saving the lives of millions of people
  
- The aims of the future larger (pivotal) study
  - ◆ Characterize the safety and efficacy of the drug
  - ◆ Motivation: drug is promising based on earlier steps
  
- The aims of the proposed pilot study
  - ◆ Evaluate feasibility of ambulatory heart monitor
  - ◆ Motivation: ensure success of the larger future study

# Designing Studies: Focus on Estimation !

---

**Do you really want a yes/no answer to your research question?**

## **RCT Example**

Research hypothesis: Drug A is better than Placebo in target population

Research objective: Estimate the relative efficacy of Drug A

Study design: RCT of Drug A vs. Placebo in a sample from the population

Expected result: Drug A will perform better than Placebo in the sample

Primary question: ~~Is the treatment effect zero?~~

Primary question: **What is the magnitude of the treatment effect?**

Martin J Gardner, Douglas G Altman (1986) “**Confidence intervals rather than p-values: Estimation rather than hypothesis testing**”

*British Medical Journal*, 292: 746-750



# Designing Studies: Focus on Estimation !

Do you really want a yes/no answer to your research question?

## Imaging Example

A radiologist plans to compare a standard imaging method (A) and a new imaging method (B) for measuring tumor volume. Agreement between volume(A) and volume(B) and the difference in Cost/Patient will be evaluated. Paired image data will be obtained. AEs will be documented.

**The proposed analysis plan is to test the following null hypothesis:**

**$H_0$ : “The correlation ( $\rho$ ) between volume(A) and volume(B) is exactly zero in the target population”.**

	Volume	Cost	AEs
Method A			
-----	-----	-----	-----
Method B			

# Designing Studies: Focus on Estimation !

**Do you really want a yes/no answer to your research question?**

## Imaging Example

A radiologist plans to compare a standard imaging method (A) and a new imaging method (B) for measuring tumor volume. Agreement between volume(A) and volume(B) and the difference in Cost/Patient will be evaluated. Paired image data will be obtained. AEs will be documented.

**The proposed analysis plan is to test the following null hypothesis:**

**$H_0$ : “The correlation ( $\rho$ ) between volume(A) and volume(B) is exactly zero in the target population”.**

### Problems

- ◆  $H_0$  is known to be false. A test of  $H_0$  will accomplish nothing.
- ◆ Wrong question: focus should be on extent of agreement.
- ◆ Focus should be on point- and interval- estimation (e.g., of  $\rho$ ).
- ◆ Correlation is just one aspects of (Bland-Altman) agreement analysis.

# Designing Studies: Focus on Estimation !

Do you really want a yes/no answer to your research question?

## Imaging Example

A radiologist plans to compare a standard imaging method (A) and a new imaging method (B) for measuring tumor volume. Agreement between volume(A) and volume(B) and the difference in Cost/Patient will be evaluated. Paired image data will be obtained. AEs will be documented.

**The proposed analysis plan is to test the following null hypothesis:**

**$H_0$ : “The correlation ( $\rho$ ) between volume(A) and volume(B) is exactly zero in the target population”.**

## Root Problem

- ◆ Opinion: Our scientific culture of over-reliance on p-values steers the investigator toward framing the main questions to have “yes / no” answers.

**The important questions are not binary; rather, they are ...**

- ◆ What is the magnitude of correlation and agreement?
- ◆ How much do we save in costs?
- ◆ How much do we gain or lose in image accuracy?
- ◆ To what extent is safety compromised (if any)?

# Designing Studies: Focus on Estimation !

---

## Essential Concept

- The focus should be on point- and interval- estimation

**Pitfalls:** research protocols frequently exhibit...

- a simplistic over-reliance on p-values together with misconceptions about p-values
- lack of focus on estimation: test procedures are mentioned, but interval-estimators and measures of precision (standard errors) are not mentioned

# Designing Studies: Focus on Estimation !

---

## Some Essential Concepts for Hypothesis Testing

- ◆ Null hypothesis: a statement about the target population
- ◆ Statistical model: all assumptions implicit or explicit
- ◆ P-value: a probability indicating how consistent the data are with the assumptions and the null hypothesis

## References

Greenland S, Senn S, et al. (2016) “Statistical tests, P-values, Confidence Intervals, and Power: a Guide to Misinterpretations.” *European J Epidemiology* 31: 337-350

Wasserstein RL, Lazar NA (2016) “The ASA's Statement on P-Values: Context, Process, and Purpose.” *The American Statistician* 70:2, 129-133.

# Designing Studies: Focus on Estimation !

## Misconceptions about hypothesis tests: P-value Fallacies

- $p\text{-value} \geq \alpha$  + Large N implies “Zero effect”
- $p\text{-value} \geq \alpha$  implies the null hypothesis is true.
- $p\text{-value} = \Pr[ H_0 \text{ is true } ]$ .
- $p\text{-value} \ll \alpha$  indicates a large important effect.
- Statistically Significant = Significant = Clinically Significant.
- It is most important to focus analysis on whether  $p\text{-value} < 0.05$ .
- The p-values are the most interesting part of the results.

All these statements are false.

# Designing Studies: Focus on Estimation !

---

**Suppose you obtained a p-value = 0.01.**

**Which of the following conclusions are true?**

- You know, if you decide to reject the null hypothesis, the probability that you are making the wrong decision.
- You have absolutely disproved the null hypothesis
- There is only a 1% probability that the null hypothesis is true.
- You have absolutely proved the alternative hypothesis
- You can deduce the probability that the alternative hypothesis is true.
- You have a reliable experimental finding, in the sense that if your experiment were repeated many times, you would obtain a significant result in 99% of trials.

**(all are false)**

# Designing Studies: Focus on Estimation !

---

Data analysis would be so easy if any of these statements were true:

- The p-values tells you whether or not an observed association or effect is real.
- If  $p > 0.05$  then that establishes that there is no association and we just say “there was no association”.
- Furthermore, to help interpret the result that  $p > 0.05$ , we can refer to our power analyses to emphasize that our study had lots of power.
- Of, if  $p > 0.05$  we can do an updated power analysis to show that the reason the p-value was large was because our study was underpowered for that test.
- On the other hand, when  $p < 0.05$  then the difference is significant and that means it is very important.

All of these sentences are false.



# Designing Studies: Focus on Estimation !

---

Data analysis would be so easy if any of these statements were true:

- If you are worrying about the possible existence of an interaction term in your regression model, look at the p-value for the interaction term; and if it is larger than 0.05 then you can be sure that no important interaction is present and the interaction term can be removed.
- A very small p-value indicates very substantial evidence of an important and strong association; because,  $p\text{-value} = \Pr[ H_0 \text{ is true } ]$ .
- You only have to look at the table of p-values to interpret the data.
- P-values are the most important results.

All of these sentences are false.

# Designing Studies: Focus on Estimation !

## References for P-value fallacies

- Altman DG, Bland JM. (1995) Absence of evidence is not evidence of absence. *British Medical Journal*, 311: 485-485.
- Hoenig JM and Heisey DM. (2001) The Abuse of Power: The Pervasive Fallacy of Power Calculations for Data Analysis, *The American Statistician*, 55(1), 19-24.
- Goodman SN, Berlin JA. (1994) The use of predicted confidence intervals when planning experiments and the misuse of power when interpreting results, *Annals of Internal Medicine*, 121, 200-206.
- Bacchetti P.(2002) "Peer review of statistics in medical research - Author's thoughts on power calculations" *British Medical Journal* 325:492-493.
- Senn SJ. (2002) "Power is indeed irrelevant in interpreting completed studies." *British Medical Journal*, 325:1304-1304.
- Bacchetti P. (2010) "Current sample size conventions: flaws, harms, and alternatives." *BMC Medicine*, 8:17
- Greenland S, Senn S, et al. (2016) "Statistical tests, P-values, Confidence Intervals, and Power: a Guide to Misinterpretations." *European J Epidemiology* 31: 337-350
- Wasserstein RL, Lazar NA (2016) "The ASA's Statement on P-Values: Context, Process, and Purpose." *The American Statistician* 70:2, 129-133.

# Designing Studies: Master Protocol Document

---

**Start the planning stage by beginning to create a master protocol document (MPD)**

- ◆ Templates are available
- ◆ Use the MPD as a repository for accumulation of ideas, information, text, literature review, and references
- ◆ For many studies, a MPD is a requirement (not optional)

# Designing Studies: Master Protocol Document

## MPD Example

0. Abbreviations
1. Purpose
2. Specific Aims
  - 2.1 Aim 1
  - 2.2 Aim 2
  - 2.3 Aim 3
3. Clinical Significance and Background
4. The Target Population of Patients
5. Recruitment of a Sample of Patients
6. Risks / Benefits for Patients
7. Study Design
  - 7.1. Treatment Design
  - 7.2. Experimental Design
  - 7.3. Measurement Design
8. Study Procedures

# Designing Studies: Master Protocol Document

( MPD Example )

9. Statistical Analysis Strategy
  - 9.1. Plans for Aim 1
  - 9.2. Plans for Aim 2
  - 9.3. Plans for Aim 3
  - 9.4. Plans for describing the sample
10. Choice of Sample Size w.r.t. Research Risk
11. Statistical Computations (who is responsible)
12. Data Management
  - 12.1 Plans for Ensuring Data Quality
  - 12.2 Plans for Ensuring Data Security
13. Randomization Protocol
14. Bibliography
15. Appendices

# Designing Studies: Master Protocol Document

---

## Even small-scale studies should have a (simple) MPD

- ◆ Relatively easy to do if the MPD is the starting point !
- ◆ Provides details of the study design, all procedures, and plans
- ◆ Serves as a working document that is updated during the study
- ◆ Avoids “*We just make it up as we go along and then try to remember what we did.*”
- ◆ Is the master source of text for the grant proposal, the IRB application, and the final publication(s).
- ◆ A grant proposal or IRB application is not a valid substitute.

# Essential concepts, Best practices, Pitfalls, Speedy IRB approval

---

- Challenges in Pilot Studies
- Challenges in Observational Studies
- Designing Studies
- **Choosing a Sample Size**
- Summary: Strategies for Speedy IRB Approval
- Appendix

# Choosing a Sample Size

---

## **If the task is to choose a sample size, N**

- We consider several possible choices for N
- For each value of N we evaluate cost, time requirements, feasibility of recruitment, anticipated precision of estimators, and anticipated power levels of test procedures (if any)
- We choose a value for N that will be satisfactory
- Judgement is always required.



# Choosing a Sample Size

---

## If N is already fixed

- For that value of N we evaluate cost, time requirements, feasibility of recruitment, anticipated precision of estimators, and anticipated power levels of test procedures (if any)
- We decide whether or not to go forward with the research.
- Judgement is always required.

# Choosing a Sample Size

---

## Valid considerations when choosing N

- aim-specific needs,
- study design constraints (e.g., N must be an even number),
- amount of research risk you are willing to take,
- amount of research risk to which others (NIH) should take,
- stage of this line of research (ranging from early to mature),
- cost in time and dollars,
- availability of subjects,
- anticipated levels of precision of estimators,
- anticipated levels of power of tests (if any).

# Choosing a Sample Size

---

## Considerations and statements that are NOT valid

- “It’s a pilot study, therefore no justification of N is needed.”
- “It’s a pilot study, therefore N should be small.”
- “A similar previous study used this N and obtained a p-value  $< \alpha$ .”
- “N = 12 is always sufficient for pilot studies.”
- Cohen’s method based on a standardized effect-size: That approach is useless; it begs the question and side-steps all the issues that should be examined.

# Choosing a Sample Size

---

## Provide a compelling rationale for the choice of N

- Small studies are not exempt from the need to state a clear and well-reasoned rationale for the number of animals or human subjects to be studied.
- For each aim, you should be able to explain in simple terms why you think  $N = 32$ , for example, is a good choice for your study. Why not 33 ?

# Choosing a Sample Size

---

## **Provide a compelling rationale for the choice of N**

- In terms of the likelihood of successfully achieving each specific aim, explain in simple language why the proposed N is a good choice.
- Make an effort to provide supporting evidence. Any calculations mentioned should be explained in sufficient detail to allow verification.

# Choosing a Sample Size

---

**Consider both: anticipated precision, anticipated power**

- For all studies, the sample size rationale should involve consideration of anticipated levels of “precision”. Why?
- For some studies, “statistical power” is irrelevant in the justification of sample size. Which ones?

# Choosing a Sample Size

---

## Choice of N requires judgement

- For a given N, each proposed hypothesis test has its own unique level of power.
- For a given N, each proposed statistical estimator has a unique level of precision.

# Choosing a Sample Size

---

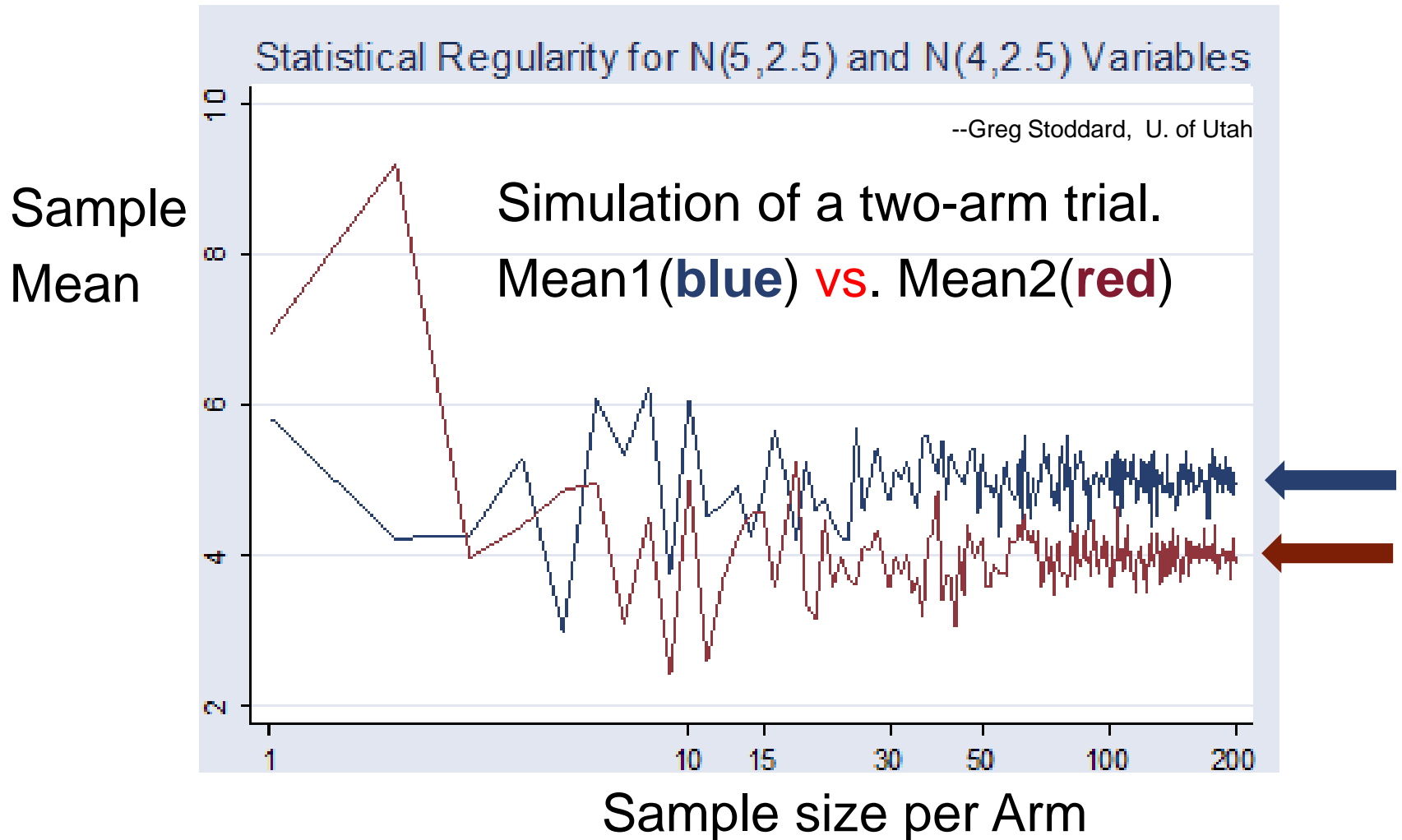
## Choice of N requires judgement

- “Sample size determination” is a misnomer. A better terminologies are
  - ◆ “power analysis”
  - ◆ “precision analysis”
  - ◆ “analysis of anticipated precision and power”.
- The target sample size (N) cannot be “determined” or calculated by a formula. It involves personal choice.



# Choosing a Sample Size: Research Risk

## Research Risk and Uncertainty Decreases with $N$



# Choosing a Sample Size: Research Risk

---

## Managing your personal “research risk”

- The analysis of anticipated power and precision is essentially an analysis of **personal research risks**.
- Ask, “How likely is it that my study will be **uninformative** and **inconclusive**?”
- The funding agency shares this risk and wants to know the answer.

# Choosing a Sample Size: Research Risk

## Personal Research Risks

- To understand the risks associated with choosing a particular sample size ( $N$ ), we have to understand that statistical hypothesis tests can be *inconclusive* and estimates can be *uninformative*.

## Uninformative Estimates

- Estimators with very wide confidence intervals (CI) are uninformative. The width of the CI is a measure of imprecision. (Precision =  $1 / SE$ .)

## Inconclusive Tests

- If the  $p$ -value  $\geq \alpha$ , then the hypothesis test is entirely inconclusive. No conclusions can be drawn or implied from the test. This is true for all choices of  $N$ .

# Choosing a Sample Size: Research Risk

## Why “inconclusive” ?

- By design, all hypothesis testing procedures are incapable of establishing that the null hypothesis is true.
- If the p-value  $\geq \alpha$ , then the hypothesis test is entirely inconclusive: it has failed to reject the null hypothesis and it is incapable of establishing that the null hypothesis is true. No conclusions can be drawn or implied from the test. This is true for all choices of  $N$ .
- Note, confidence intervals (interval estimates) are always informative to some degree, and can be highly informative (narrow) even when the test is inconclusive.

# Choosing a Sample Size: Research Risk

---

## Why “uninformative” ?

- Precision = ( 1 / SE )
- $W$  = expected half-width of a 95% confidence interval
- $W$  = “the margin of error”
- $W \approx 2$  or 3 times the standard error (SE)

# Choosing a Sample Size: Research Risk

## Appropriate use of “**inconclusive**”

“Outcome  $Y$  was associated with  $X_1$  in the study cohort and the null hypothesis “no association in the target population” was rejected (Odds Ratio = 2.2 [1.2, 3.9],  $p = 0.009$ ).”

“The statistical hypothesis test of association between  $Y$  and  $X_2$  was inconclusive (OR = 1.3 [0.7, 2.6],  $p = 0.31$ ); sufficient evidence was not available to establish that there is no association in the target population. The OR estimate and 95% confidence interval suggest that there may be a small or substantial association.”

Avoid the use of easy-to-say but misleading phrases such as

- there was no association
- we saw no evidence of association
- there was no statistical difference
- association was not detected

# Choosing a Sample Size: Research Risk

## Appropriate use of “inconclusive”

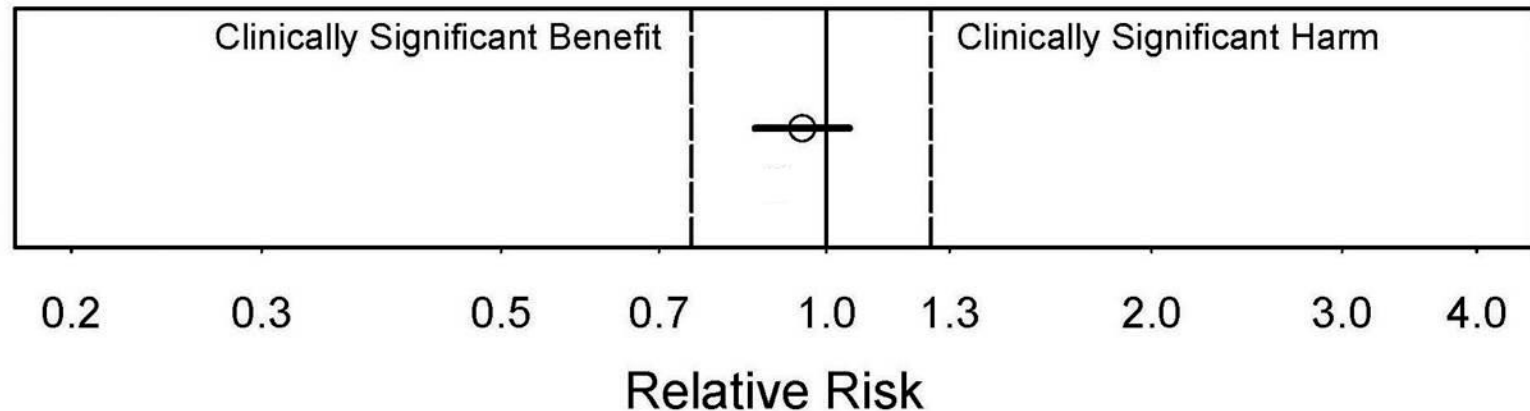
“The statistical hypothesis test of association between  $Y$  and  $X_2$  was inconclusive (OR = 1.3 [0.7, 2.6],  $p = 0.31$ ); sufficient evidence was not available to establish that there is no association in the target population. The OR estimate and 95% confidence interval suggest that there may be a small or substantial association.”

Question: “That makes it sound like there might be an association when in fact the result was negative; isn’t that misleading?”

Answer: “That is the point: our study has failed to establish that the null hypothesis is true (i.e., OR=1.0 in the target population). The point estimate and confidence interval suggest that there may be notable association in the population. Association was observed in the sample and we have not established that there is no association in the population. It would be incorrect and misleading to say ‘*there was no association*’.”

# Choosing a Sample Size: Research Risk

## Example: Relative Risk

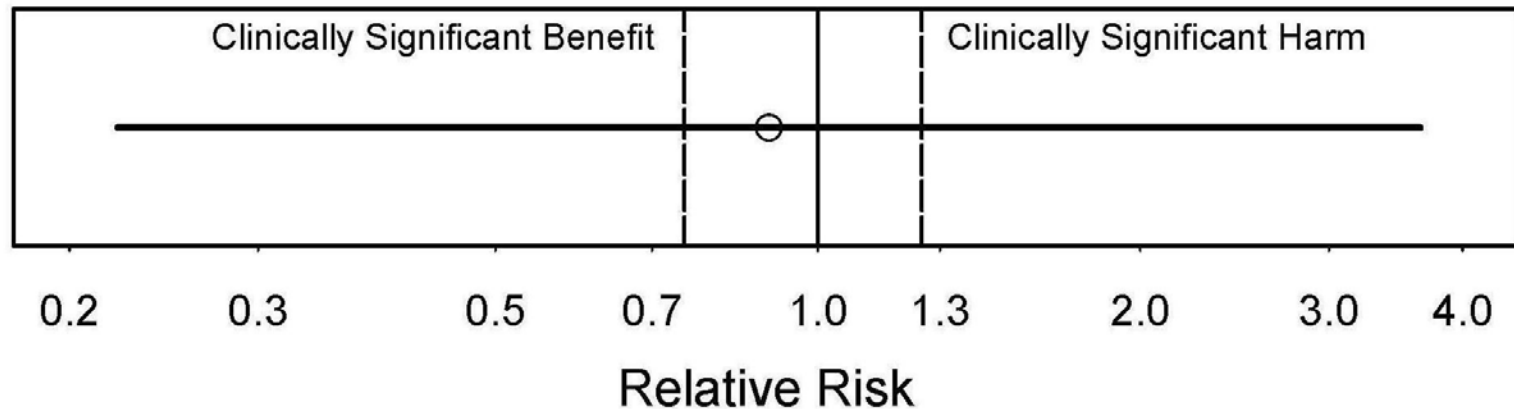


We found strong evidence against any substantial harm or benefit.



# Choosing a Sample Size: Research Risk

## Example: Relative Risk



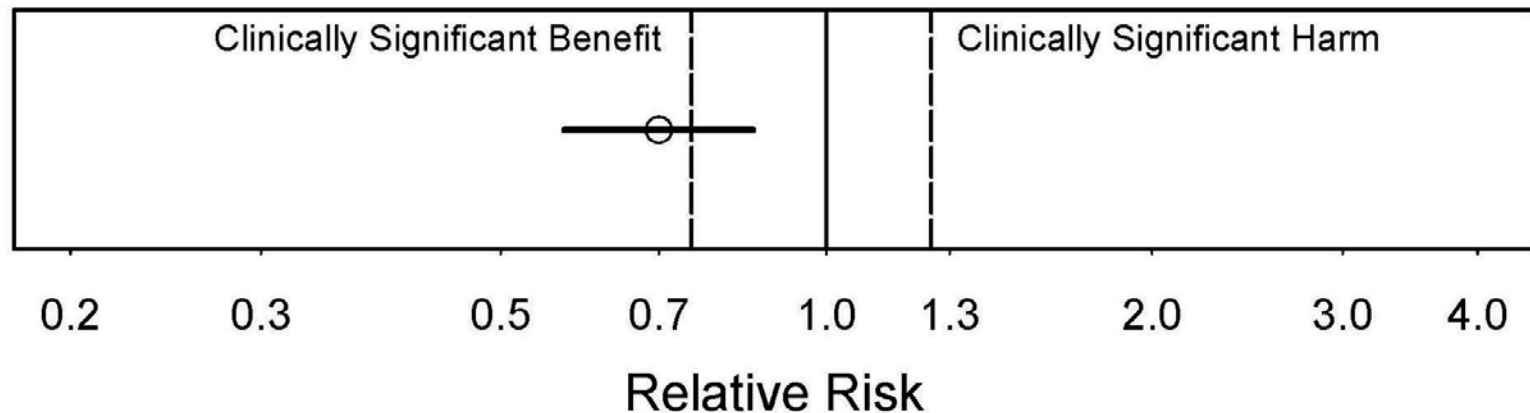
No conclusions possible due to very wide CI  
(Lack of precision)

-- Peter Bacchetti

# Choosing a Sample Size: Research Risk

## Example: Relative Risk

- Let's focus on the magnitude, direction, and precision of effects
- Requires discussion of magnitudes that are “clinically significant”



Strong evidence of benefit (statistically significant)

Substantial benefit appears likely, but CI too wide to rule out clinically unimportant benefit

-- Peter Bacchetti

# Choosing a Sample Size: Research Risk

---

- Avoid over-reliance on p-values
- Avoid misinterpretation of p-values
- Focus on point estimates and confidence intervals
- The statistical estimates and their confidence intervals convey much more information than p-values.
- Precision should be an important consideration when choosing a sample size.

# Choosing a Sample Size: Research Risk

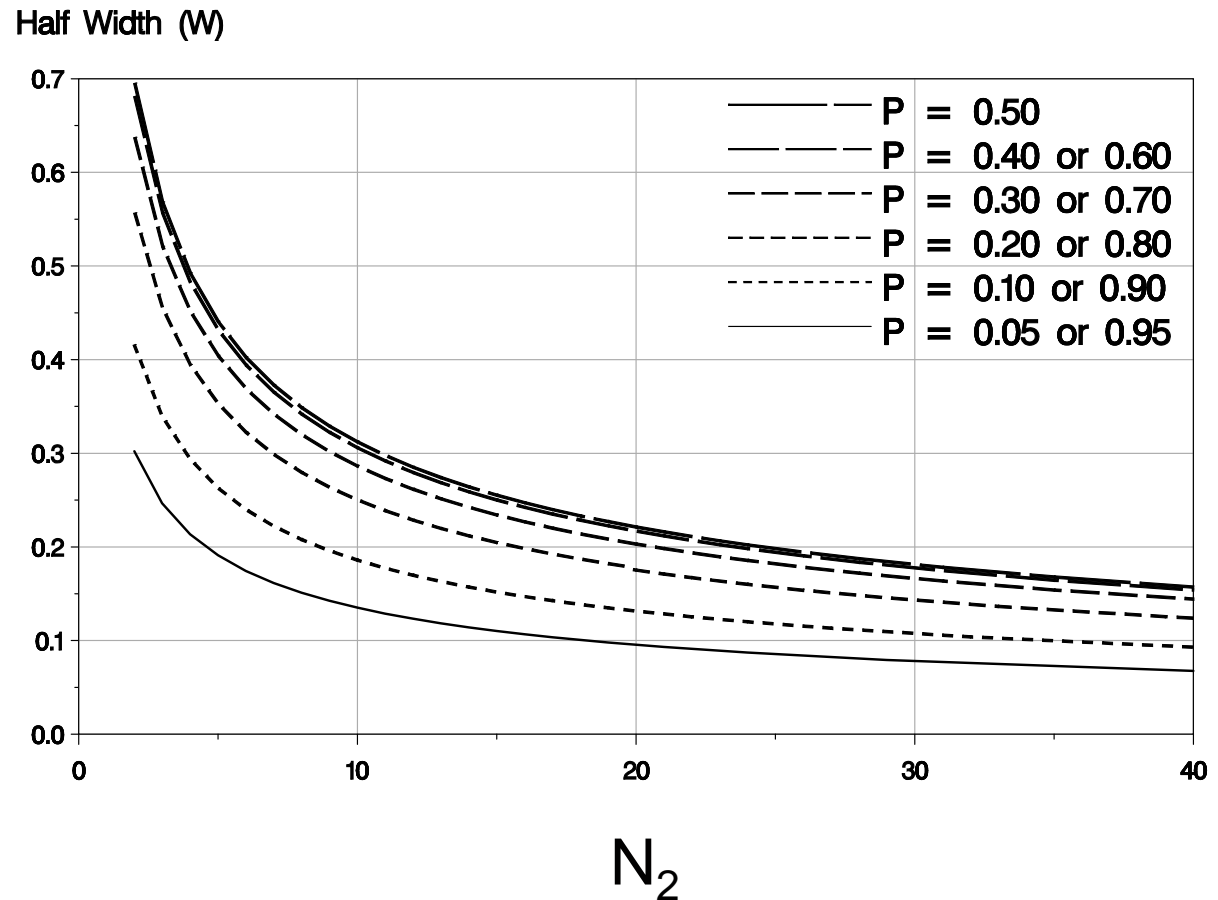
---

- Frequently, a sample size may provide a “satisfactory” level of anticipated power for a test, while NOT providing a “satisfactory” level of precision for the important statistical estimates of interest.
- Precision should be an important consideration when choosing a sample size.

# Choosing a Sample Size: Research Risk

**Precision:** Anticipated width of a 95% CI for a proportion,  $P$

If the observed proportion is  $P$ , the corresponding 95% confidence interval (CI) will be  $P \pm W$  in which  $W$  is computed as 1.96 times the standard error of  $P$ .



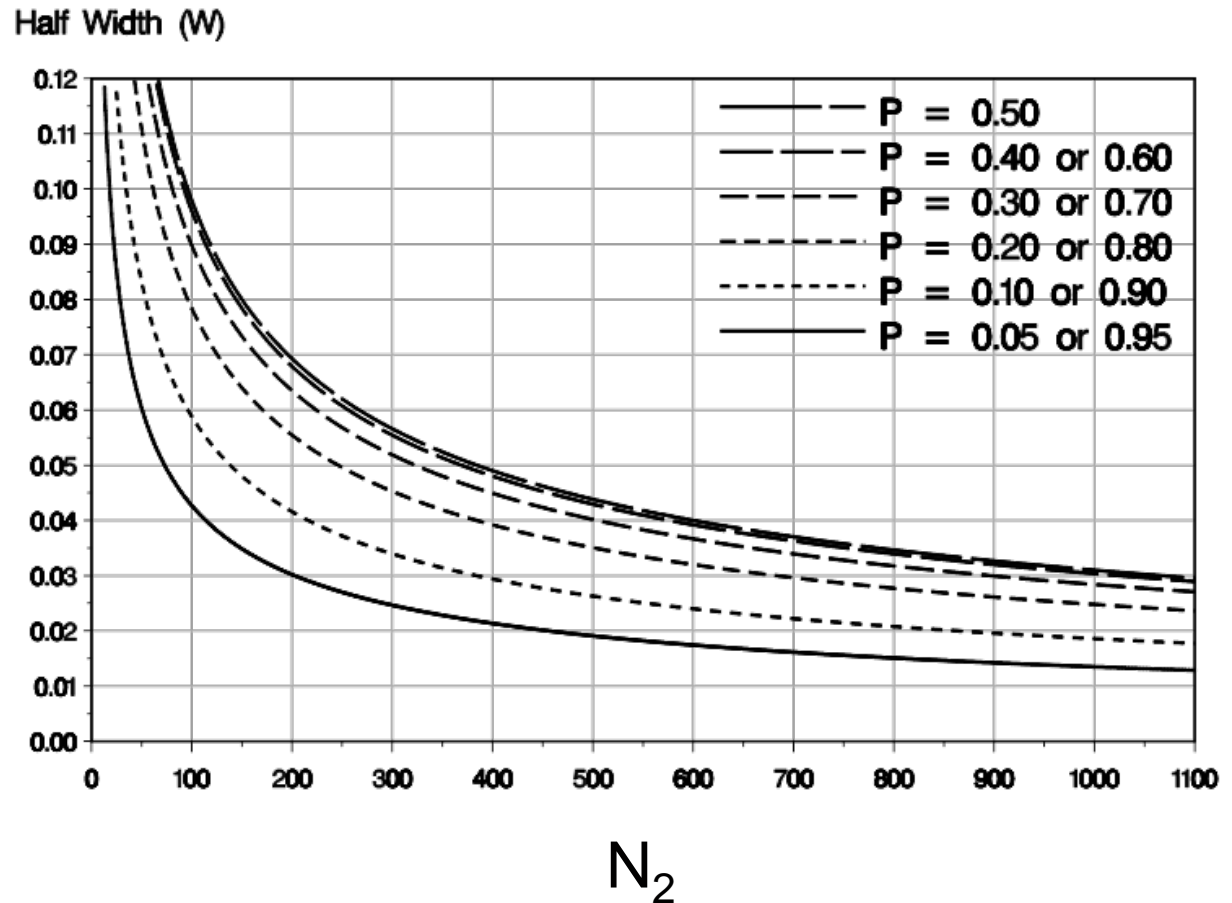
# Choosing a Sample Size: Research Risk

## Precision: Anticipated width of a 95% CI for a proportion, P

If the observed proportion is  $P$ , the corresponding 95% confidence interval (CI) will be  $P \pm W$  in which  $W$  is computed as 1.96 times the standard error of  $P$ .

If  $N_2=100$  and  $P=0.50$  then  $W=0.10$  and the CI is  $[0.40, 0.60]$ .

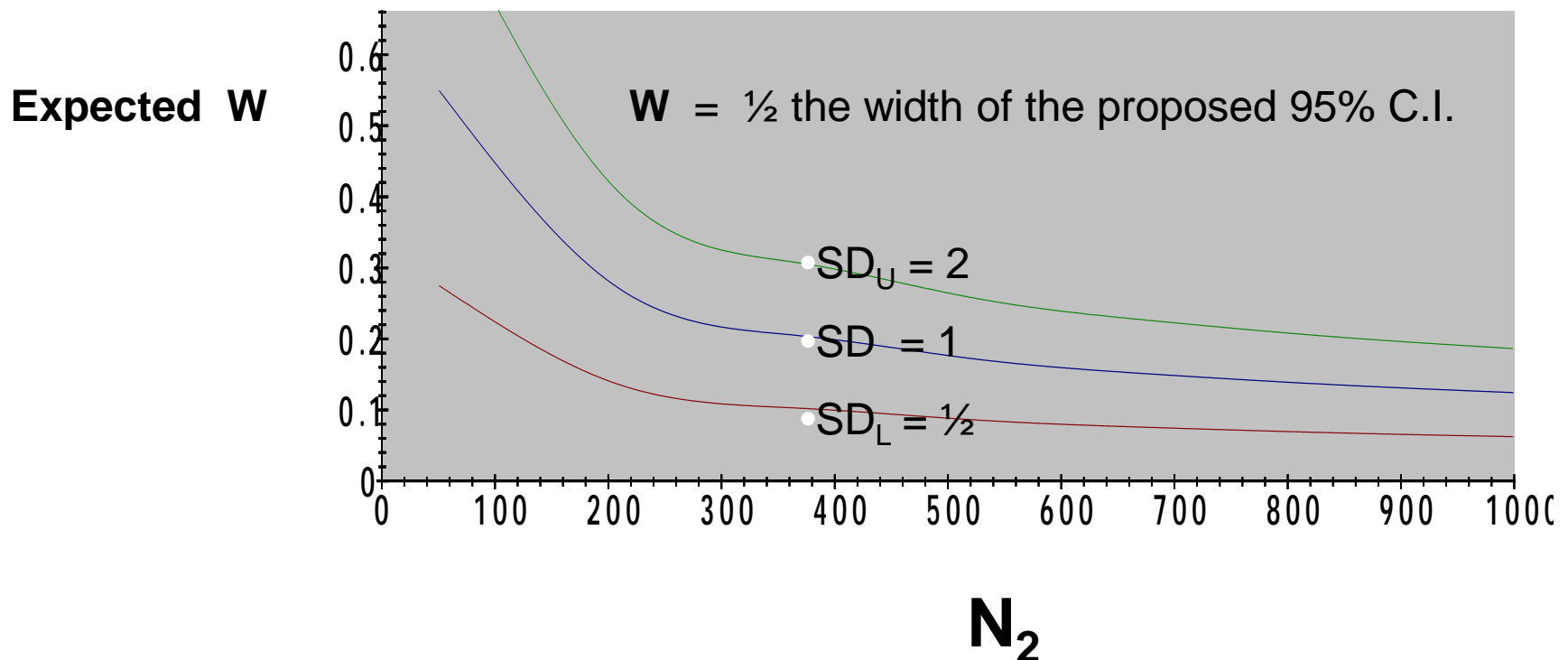
If  $N_2=1067$ ,  $P=0.50$  then  $W=0.03$  and the CI is  $[0.47, 0.53]$ .



# Choosing a Sample Size: Research Risk

## **Precision:** Anticipated width of a 95% CI for a Mean

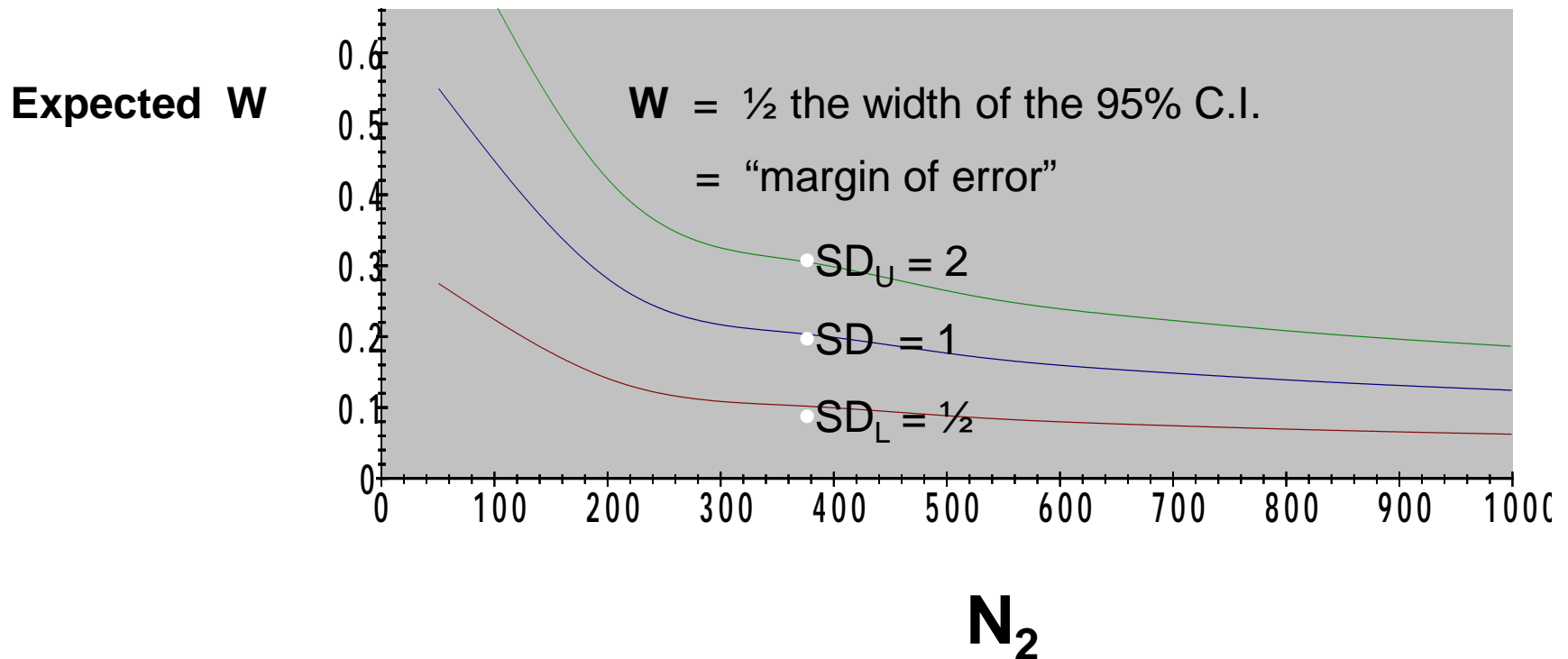
For a given proposed  $N_2$ , the expected width of a proposed confidence interval is a function of the population standard deviation ( $\sigma$ ). But  $\sigma$  is unknowable. We use candidate values; e.g., a confidence interval for  $\sigma$  based on a previous sample of size  $N_1$  provides a point estimate (**SD**) as well as a plausible range [**SD<sub>L</sub>**  $\leq \sigma \leq$  **SD<sub>U</sub>**]



# Choosing a Sample Size: Research Risk

## **Precision:** Anticipated width of a 95% CI for a Mean

There is very roughly a 50% chance that the observed C.I. will be wider (or narrower) than the expected width. An alternative is to plan for the C.I. to be narrower than some preselected width with high probability.





# Choosing a Sample Size: Research Risk

## Precision: for a Mean

Confidence Interval's Property

$$\Pr[ \mathbf{SD}_L \leq \sigma \leq \mathbf{SD}_U ] = 80\%$$

Observed based on  $\mathbf{N}_1 \dots$

$$\mathbf{SD}_L = 5.0$$

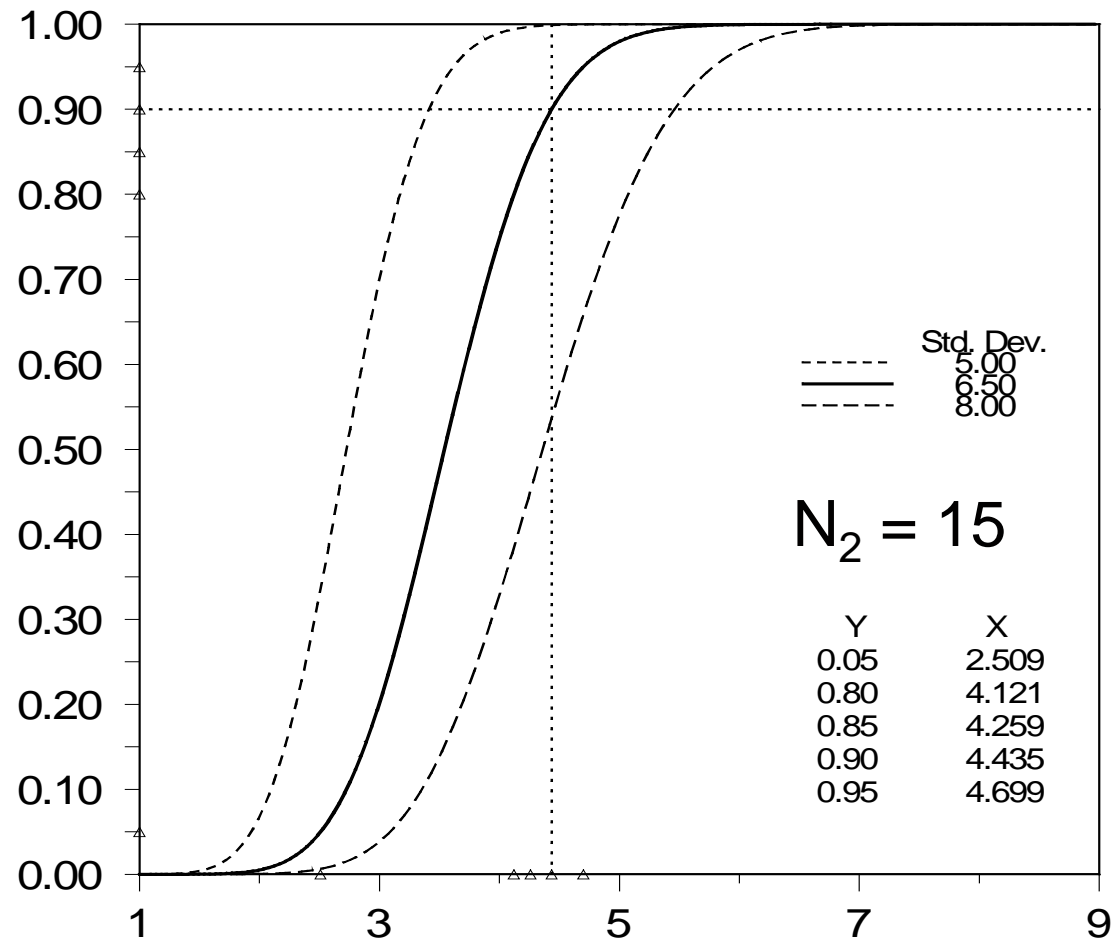
$$\mathbf{SD} = 6.5$$

$$\mathbf{SD}_U = 8.0$$

$W$  = half-width of the C.I. that will be observed in the future study.

If population parameter  $\sigma$  happens to equal 6.50, then there is a 90% chance that  $W$  will be  $\leq 4.435$

$$\Pr[ W \leq w \mid \sigma, N_2 ]$$



**W**

# Choosing a Sample Size: Research Risk

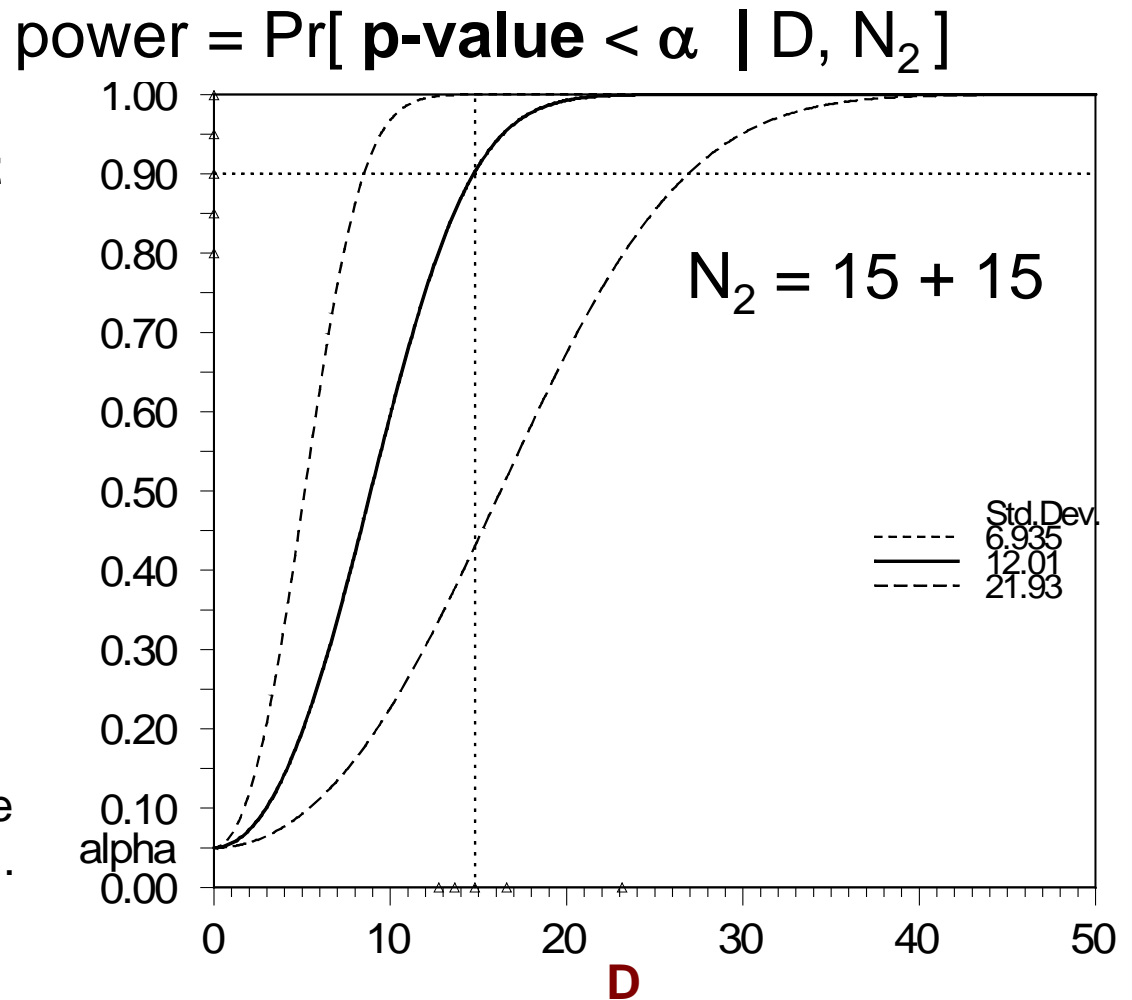
**Power:** for a hypothesis test

## Power Curves

The (null) hypothesis tested is  $H_0$  “ $D=0$  in the target population”

For all choices of  $N_2$ , if  $H_0$  is true then power =  $\alpha$ .

In this example, the wide C.I. for the standard deviation yields a wide C.I. for the estimate of power.



Population mean difference between Regimens

# Choosing a Sample Size: Research Risk

Confidence Interval's property

$$\Pr[ \mathbf{SD}_L \leq \sigma \leq \mathbf{SD}_U ] = 80\%$$

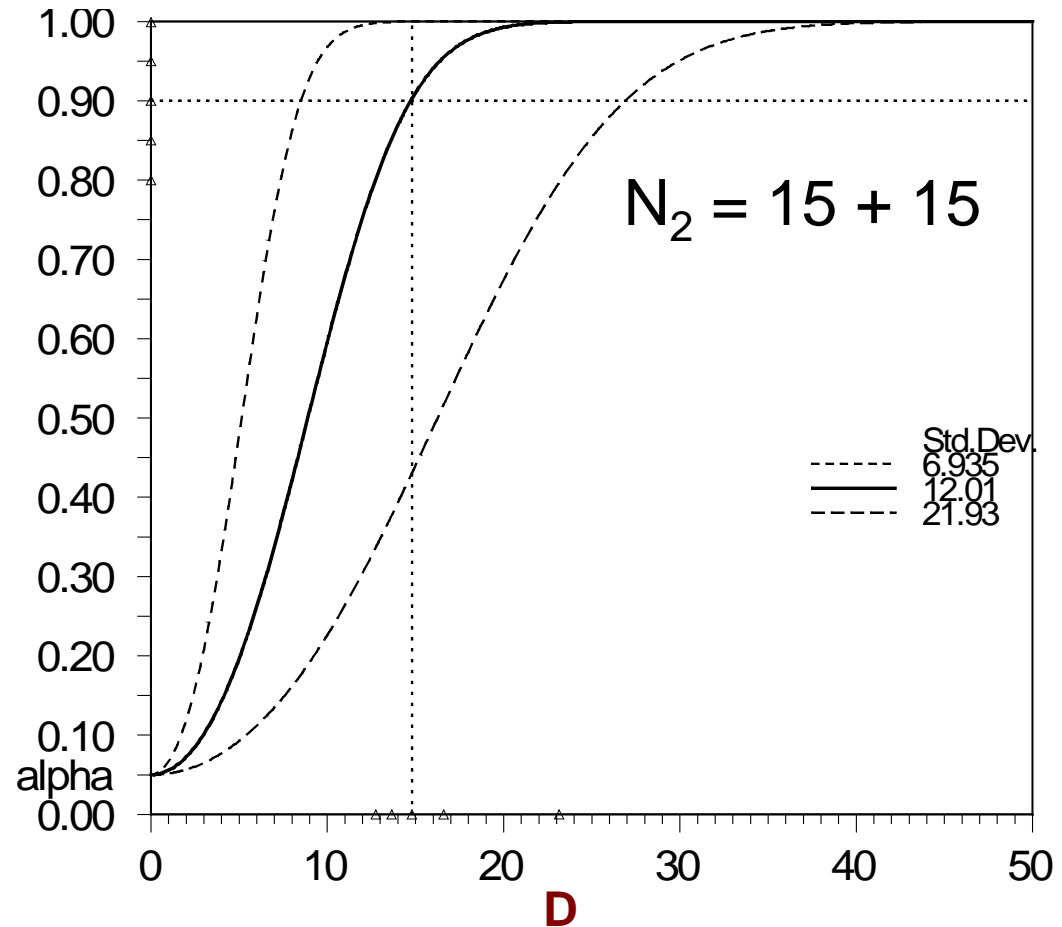
Observed based on  $\mathbf{N}_1 \dots$

$$\mathbf{SD}_L = 6.9$$

$$\mathbf{SD} = 12.0$$

$$\mathbf{SD}_U = 21.9$$

$$\text{power} = \Pr[ \text{p-value} < \alpha \mid D, N_2 ]$$



Population mean difference between Regimens

---

## References

Browne RH (1995). On the Use of a Pilot Sample for Sample Size Determination. *Statistics in Medicine* 14: 1933-1940.

Taylor DJ and Muller KE (1995). Computing Confidence Bounds for Power and Sample Size of the General Linear Univariate Model. *The American Statistician* 49: 43-47.

Recommended resource: *[samplesizeshop.org](http://samplesizeshop.org)*

# Essential concepts, Best practices, Pitfalls, Speedy IRB approval

---

- Challenges in Pilot Studies
- Challenges in Observational Studies
- Designing Studies
- Choosing a Sample Size
- **Summary: Strategies for Speedy IRB Approval**
- Appendix

# SUMMARY: STRATEGIES FOR SPEEDY IRB APPROVAL



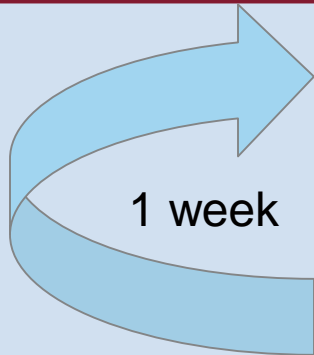
# Scientific Review

- All clinical research at UNC-CH involving greater than minimal risk is reviewed by the full board of the IRB and must (first) undergo scientific review. Some exceptions apply.
  - ◆ **Scientific Review Committee (SRC)** in the Office of Clinical Trials (OCT)
  - ◆ Protocol Review Committee (PRC) in the L.C.C.C.
  - ◆ other agencies providing scientific review
- IRB is then free to focus on ethical concerns (stipulations)
- SRC reviews and IRB reviews provide constructive criticism

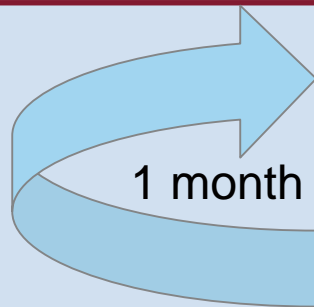
# SRC & IRB Review: August 2016 – Present

**SRC** -- Scientific Review Committee

**MPD** -- Master Protocol Document



1. Start IRB Application and Create **MPD**
2. Submit **MPD** to the **SRC** for scientific review
3. Receive **SRC** reviews (clinical and statistical)
4. Revise **MPD** and resubmit to resolve concerns
5. **SRC** (re-)Review and **MPD** provided to IRB



1. Complete the IRB Application
2. Receive IRB stipulations (focused on ethics)
3. Revise and resubmit
4. IRB approval / decision



# Speedy Approval: Which Studies?

- Delays in approval are more typical for ...
  - ◆ Local investigator-initiated protocols
  - ◆ Pilot studies by new investigators
  
- Speedier approval is typical for ...
  - ◆ Studies by research teams that include biostatisticians
  - ◆ Research network studies (e.g., ACTG, TDN-CF)
  - ◆ Industry-sponsored studies
  - ◆ Studies that involve a CRO or coordinating center
  - ◆ Some NIH RFA programs and FDA-funded studies



# Speedy Approval: Which Studies?

- Few stipulations: **Why?**
  - ◆ Use of best practices
  - ◆ Research teams include professionals in multiple disciplines
    - Biostatistics,
    - Regulatory Affairs,
    - Data Management,
    - Systems Programming, etc.
  - ◆ Benefit of multidisciplinary development of the protocol
  - ◆ Funding for collaborative input and support
  - ◆ Previous cycles of review (e.g. FDA) and refinement



# Strategic Topics

1. Consulting / Collaborating Early with Supportive Professionals
2. Master Protocol Document
3. Answering Questions in the Online IRB Application
4. Addressing Stipulations
5. Data Management Plans for Data Quality
6. Alignment of Aims with Design and Analysis
7. **Study Design**
8. Statistical Analysis Plans
9. Choice of Sample Size w.r.t. Research Risk
10. Inclusion of Essential Expertise on the Research Team



# #7. Study Design

## How many ways could this go wrong?

- 7.1 Specific aims
- 7.2 Pilot study protocol design
- 7.3 Design features
- 7.4 Randomization protocols
- 7.5 Blinding
- 7.6 Variables well-defined
- 7.7 Stratification / cohorts
- 7.8 Matching



# #1. Early Collaboration

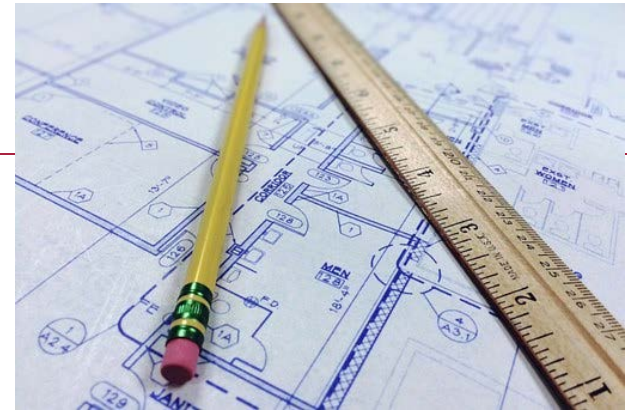
**Consult early with professionals in Biostatistics, Regulatory, Data Management, and other specializations**

If your research team does not already include statistical expertise, contact the TraCS Biostatistics Core in the earliest planning stage of any new research, months in advance of SRC & IRB review.



The NC TraCS Institute provides resources, guidance, consultations, and helpful core units; e.g., the Regulatory Core, Biostatistics Core, and Biomedical Informatics Core.

## #2. Master Protocol Document



**Strengthen the research by creating and maintaining a Master Protocol Document**

A MPD is a living document that completely specifies all details of the research project.

The MPD contains summary statements as well as intricate details of procedures and methods.

During earliest planning stages, the evolution of the protocol is reflected in draft revisions of the MPD.

During execution of the study, the working version of the MPD is updated / expanded to capture new information.

## #2. Master Protocol Document

**Strengthen the research by creating and maintaining a Master Protocol Document**



A research *proposal* is not a valid substitute for the MPD.  
An IRB application is not a valid substitute for the MPD.  
Those documents do not provide sufficient detail for the study.

If an IND / IDE is required, a MPD is required by FDA.

MPD is critical for coordination of multi-center studies.

All human studies can benefit from use of a MPD that is as brief & simple or as long & complex as the study it represents.

# #3. Answer the Questions

---

**Carefully answer all the questions in the IRB application.  
Do not be tempted to omit an answer or guess what is needed.**

In the online IRB application: investigators often need guidance.

If the answer to a question is not obvious, or if two questions seem redundant, seek clarification and assistance by contacting the IRB office, or UNC TraCS Institute's helpful personnel, or other knowledgeable experts.

Contact the TraCS Biostatistics Core for help with questions about statistical considerations.





# #4. Address SRC Concerns and IRB Stipulations

---

**Carefully address all review comments.**

**Seek expert assistance with the response and revision.**

If a review comment seems unclear, or an appropriate response is not obvious, contact knowledgeable resources for assistance; e.g., TraCS cores, SRC coordinator at OCT, or in the IRB office.



# #5. Data Management Plans

**Present a data management plan for ensuring data quality in addition to ensuring data security and confidentiality**

Data Quality: How efficiently the data serves the purpose of achieving the specific aims.

High quality avoids “Garbage In, Garbage Out” or the even worse... “Garbage In, Gospel Out”.

Every study generates a database --a collection of information (data files and documentation) used to achieve the specific aims.

Simple studies yield simple databases.

Collaborate early on writing the plans.



# #5. Data Management Plans

---

## Aspects of Data Quality

- Collecting the “right” data?
- Completeness of the data.
- Coding of the reasons for missing values.
- Times of measurements are recorded
- Accuracy of the data ...
  - ◆ Accuracy during data collection and data entry
  - ◆ Error detection during data monitoring and validation
  - ◆ Error detection during data cleaning



# #5. Data Management Plans

## Aspects of Data Quality

- Documentation
  - ◆ Codebook (units, ranges, definitions)
  - ◆ Audit trail (log of what changes, when, who)
  - ◆ Event journal
  - ◆ Version numbers on CRFs and surveys
  - ◆ Internal & external documentation of computer code/macros
  
- Adherence to a codebook (data dictionary)
  - ◆ Good:  $\text{gender} \in \{M, F\}$
  - ◆ Bad:  $\text{gender} \in \{M, F, m, f, \text{Male}, \text{Female}, \text{M-dropout}\}$



# #5. Data Management Plans

## Aspects of Data Quality

- Data Integrity
  - ◆ Protection from unintended changes to the data
  - ◆ Protection from corruption due to human error, hardware failure, software bugs, malicious intent
  
- Ensuring Data Integrity
  - ◆ Authentication (login with password)
  - ◆ Authorizations (user enters new data, but cannot edit old data)
  - ◆ Backup and archival
  - ◆ Validation of software
  - ◆ Electronic signatures (weak)
  - ◆ Digital signatures (strong)



# #5. Data Management Plans

## Aspects of Data Quality

- Coordination
  - ◆ Training and monitoring of personnel collecting data
  - ◆ Written procedures manual for data collection
  
- Pilot testing operations (this is unavoidable)
  - ◆ Validation of systems, software, macros, functions, programs
  - ◆ Verifying adequacy of codebook specifications
  
- For some studies: compliance with regulatory standards
  - ◆ HIPAA compliance
  - ◆ FDA 21 CFR Part 11 compliance



# #5. Data Management Plans



## Software for Data Capture

- Features that facilitate best practices
  - ◆ uses a data dictionary (codebook)
  - ◆ easy-to-use forms for data entry
  - ◆ facilitates data monitoring and data editing
  - ◆ ensures integrity of the data
  - ◆ HIPAA compliant for handling PHI data
  - ◆ web-based and remotely accessible
  - ◆ audit trail is automatic (what changes made by whom, when)
  - ◆ authentication (login with password)
  - ◆ authorization for role-based access (e.g., entry but not export)
  - ◆ compatible with SAS, SPSS, R, Stata



# #5. Data Management Plans



## Software for Data Capture / Data Entry

- REDCap – “Research Electronic Data Capture”
  - ◆ best data capture software at UNC-CH
  - ◆ available at no cost to UNC-CH investigators
  
- CDART – from TraCS with CSCC for multicenter
- EPINFO – free from CDC, but effort needed to set it up
- Access – solves some of the problems with Excel
- Access + SQLserver – requires programming
- OpenClinica – expensive
- ORACLE Clinical – expensive





# #5. Data Management Plans



## Software for Data Capture / Data Entry

- MS Excel is a poor choice for biomedical research data
  - ◆ Not designed for data entry and data management
  - ◆ High risk of loss of data or corruption of the data
  - ◆ Facilitates very poor data management practices
  - ◆ Inadequate for data that requires HIPAA compliance
  - ◆ No capacity to store metadata with data values
  - ◆ No audit trail
  - ◆ No skip patterns. Data entry forms difficult to set up
  - ◆ No authentication. No authorization.
  - ◆ Sorting of columns is a huge issue
- MS Access is a slight improvement



## #6. Alignment of Aims, Design, Analysis

---

**For each aim present an aim-specific statistical analysis strategy that is adequately detailed and appropriate for the data.**

If data are longitudinal and aims require longitudinal analysis, then the analysis strategy should rely on methods appropriate for longitudinal analysis (e.g., repeated-measures ANOVA).

For each specific aim there should be a statistical analysis plan.

On the research team, include personnel with the expertise required to perform the statistical computations and interpretive analyses needed for each aim.



# #7. Study Design

## How many ways could this go wrong?

- 7.1 Specific aims
- 7.2 Pilot study protocol design
- 7.3 Design features
- 7.4 Randomization protocols
- 7.5 Blinding
- 7.6 Variables well-defined
- 7.7 Stratification / cohorts
- 7.8 Matching



# #7.1 Study Design: Specific Aims

**All aims should be stated clearly and realistically.**

- Common pitfalls
  - ◆ None of the aims are explicitly stated
  - ◆ Some aims are stated, others go unrecognized
  - ◆ Unrealistic statement of aims (e.g., not “To prove Drug A is completely safe”)
  - ◆ In statement of aims, confusing the aims of the pilot study with the aims of the pivotal study.
    - In the example from 1998, that uncontrolled study of N=10 patients was never capable of “Determining the efficacy of Drug A”



# #7.1 Study Design: Specific Aims

**R C T Example:** *P* = placebo, *A* = dosage 1, *B* = dosage 2

Aim1. For *A* and for *B*, evaluate relative safety during treatment

Aim2. For *A* and for *B*, estimate PK profile on day 7 of Treatment.

Aim3. For *A* and for *B*, estimate treatment success rate at 3 mon.

Aim4. Compare *A* vs. *B* in terms of relative efficacy.

Aim5. Explore data to generate hypotheses (predictors of efficacy).

Aim6. Estimate the rate of recruitment and the drop-out rate.

Aim7. Estimate components of variance (use to plan next study).

4-7 are preparatory for a pivotal study of drug vs placebo.



## #7.2 Study Design: Pilot Study

**If planning any kind of pilot study, consult early with professionals in biostatistics, data management, and regulatory.**

- Moore CG, Carter RE, Nietert PJ and Stewart PW (2011). “Recommendations for planning pilot studies in clinical and translational research.” *Clinical & Translational Science*, 4(5), 332-337.
- Many pitfalls to avoid during planning stages and when applying for IRB approval.



# #7.3 Study Design: Features & Rationale

**Provide a rationale for the study design.**

- Design Features
  - ◆ Controlled / Uncontrolled,
  - ◆ Placebo / No Treatment,
  - ◆ Cross-sectional / Longitudinal,
  - ◆ Randomized / Not,
  - ◆ Observational / Experimental,
  - ◆ Prospective / Retrospective,
  - ◆ Interim Analyses / No interim Analyses.
  
- Example: Why should this be an uncontrolled study?



## #7.3 Study Design: Features Rationale

Provide a rationale for the study design.

- **Treatment design:** specification of conditions/treatments
  - ◆ Why this dosage?
  - ◆ Why no placebo?
  - ◆ Why open-label?
  
- **Experimental design:** how treatments are assigned to subjects
  - ◆ Why is randomization not used?
  - ◆ Why is the observational design necessary?
  
- **Measurement design:** what measurements and when
  - ◆ Why exactly that many longitudinal measures?





## #7.4 Study Design: Randomization

---

**Specify whether / how randomization will be achieved and who will perform the computations.**

**Example text:** “For each gender a randomization table will be computed by the data management personnel using a method of permuted blocks of size 2 and 4. Only patients verified as being eligible for enrollment will be randomized. The treatment assignments will be concealed from the personnel enrolling patients. The randomization tables will be used by the Investigational Drug Service (IDS) in their sequential provision of treatment regimens (identified only by *Subject\_ID* labeling.) No other personnel will have access to the randomization schedule. Regardless of drop-outs, each new subject will be assigned by randomization according to the randomization tables. Further details are provided in the MPD.)”

## #7.5 Study Design: Blinding

**Specify a plan for blinding (who, when) and provide a rationale for the proposed approach.**

Who will be blind to the patient's treatment/conditions, and when will they become unblinded?

Who will have access to the data, and when will they have it?

If interim computations are performed, who will have access to the results and will they be able to influence decisions to continue / stop the study ?



## #7.6 Study Design: Define Variables

**For each aim, the outcome variables and other measures of interest should be well-defined.**

The units of the variables as will be used in statistical analysis should be clearly defined.

Example: “serum viral burden ( $\log_{10}$  RNA copies/mL) measured at 0, 3, 6 months post-treatment.”

Decisions about whether to transform the scale (e.g., sqrt,  $\log_{10}$ ) should be made prior to recruitment of subjects.



# #7.7 Study Design: Stratification

**If stratification is used, the details should be clearly specified, and the role of stratification in statistical analysis should be explained and justified.**

- For example, used in stratified randomization and in statistical analysis methods for stratified data.
  
- Common pitfalls
  - ◆ It is unclear how or whether stratification will occur
  - ◆ Role of stratification not explained
  - ◆ Rationale for stratification not provided



## #7.8 Study Design: Matching

If matching is used, the details should be clearly specified, and the role of matching in statistical analysis should be explained and justified.

- Common pitfalls
  - ◆ It is unclear whether matching will occur
  - ◆ No details of how the matching will be performed
  - ◆ Role of matching not explained
  - ◆ Rationale for matching not provided
  
- Having equal numbers of males and females in the treatment groups is not “matching”



# Strategic Topics

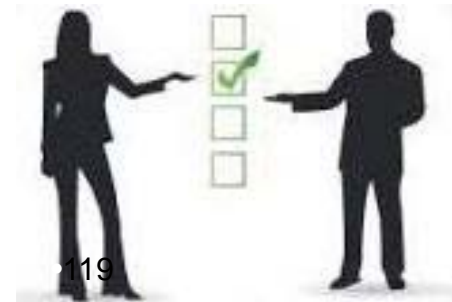
1. Consulting / Collaborating Early with Supportive Professionals
2. Master Protocol Document
3. Answering Questions in the Online IRB Application
4. Addressing Stipulations
5. Data Management Plans for Data Quality
6. Alignment of Aims with Design and Analysis
7. Study Design
8. **Statistical Analysis Plans**
9. **Choice of Sample Size w.r.t. Research Risk**
10. **Inclusion of Essential Expertise on the Research Team**



# #8. Statistical Analysis Plans

**For each aim, an appropriate statistical analysis strategy should be explained in complete detail.**

- Frequent causes of concerns raised
  - ◆ No plans presented.
  - ◆ No plans presented; instead, assays are described, measures defined, or outcome variables explained.
  - ◆ Some plans, but ambiguous and lacking in sufficient detail.
  - ◆ Plans are incomplete; some specific aims not addressed.
  - ◆ Plans are incomplete; issues not addressed (e.g., missing data)



# #8. Statistical Analysis Plans

**For each aim, an appropriate statistical analysis strategy should be explained in complete detail.**

- Frequent causes of concerns raised
  - ◆ Presented plans are inappropriate (e.g., wrong method)
  - ◆ Statistical misconceptions evident in the narrative
  - ◆ Incorrect use of statistical terminology
  - ◆ No strategy presented; only a list of methods is mentioned.  
“We will use t-tests and chi-square tests.”  
analogous to “We will use pills and needles and tubes.”





# #8. Statistical Analysis Plans

**For each aim, an appropriate statistical analysis strategy should be explained in complete detail.**

- Frequent causes of concerns raised
  - ◆ Over-reliance on p-values, lack of focus on magnitudes of estimates and their confidence intervals.
  - ◆ Decisions about whether to transform the scale (e.g., sqrt,  $\log_{10}$ ) should be made prior to recruitment of subjects
  - ◆ No analysis strategy for coping with incomplete data: missing values, censored assay values, etc.



# #8. Statistical Analysis Plans

**For each aim, an appropriate statistical analysis strategy should be explained in complete detail.**

- Frequent causes of concerns raised
  - ◆ Unclear if interim analyses will be performed.
  - ◆ Plans for interim analysis are ambiguous or are missing.
  - ◆ No analysis strategy for avoiding inflation of Type I error rates due to multiplicity of hypothesis testing.
  - ◆ Or, no justification of why such a strategy is not needed.



# #8. Statistical Analysis Plans

**For each aim, an appropriate statistical analysis strategy should be explained in complete detail.**

## ■ **For best results...**

- ◆ Begin collaboration with professionals in biostatistics in the earliest stage of planning (months in advance of deadlines).
- ◆ The research team should include professional biostatistician(s) or other co-investigator(s) with statistical expertise.
- ◆ Analysis plans and other statistical sections of the protocol should be drafted by one of those co-investigators.
- ◆ Start with a master protocol document (MPD).  
Copy from it to create grant apps and IRB apps.



# #8. Statistical Analysis Plans

For each aim, an appropriate statistical analysis strategy should be explained in complete detail.

## For best results...

- Inferential analyses should be completely specified prior to collection of data (*a priori*).
- The analysis plans should mention use of sensitivity analyses to evaluate the robustness of the study's main results to reasonable perturbations of the statistical methods and assumptions used.
  - ◆ While there are always competing statistical methods from which to choose, to help ensure reproducibility of research the main results should be obtained using a single choice of methods that is specified a priori; thus, uncertainty about the optimal choice of methods and assumptions is best handled by relegating competing approaches to an important role in the domain of sensitivity analyses.
  - ◆ Results of the sensitivity analyses should only be used to guide trust in the main results.



# #9. Choice of Sample Size

**All human research proposals should present a compelling rationale for the choice of sample size (N).**

- In terms of the likelihood of achieving each aim, explain in simple language why the proposed N is a good choice. Provide supporting evidence.
- Aims may or may not be achieved. The results depend on which patients happen to be recruited, how measurement errors occur, etc. Generally, larger N reduces the risk.
- Risk: the likelihood that the study's results will be uninformative, inconclusive, not useful and ... not published.



# #9. Choice of Sample Size

## Valid considerations for choosing the sample size:

- How much risk ...for the research team? ...for funding agency?
- The relative importance (priority) of each aim
- The sample-size needs of each aim
- Stage of research (small N for “first time in humans”)
- Costs in time and money
- Availability of eligible research subjects
- Anticipated % of patients who will drop-out
- Anticipated % of patients with complete data
- Anticipated levels of precision of estimators
- Anticipated levels of power for the tests
- Discussions of “clinically significant” magnitudes



# #10. Essential Expertise on the Team

**Identify personnel who will be responsible for all aspects of the study; especially... study coordination, computations for data management, statistical computations for data analysis, interpretive analysis of the results, and collaboration on statistical aspects of manuscript preparation.**



# Essential concepts, Best practices, Pitfalls, Speedy IRB approval

---

- Challenges in Pilot Studies
- Challenges in Observational Studies
- Designing Studies
- Choosing a Sample Size
- Summary: Strategies for Speedy IRB Approval
- **Appendix**



## Appendix

### A simulation of the impact of using unreliable input in a sample-size analysis



# Pilot study ( $N_1$ ) to estimate SD for future study

---

## Suppose we have a study with a mix of aims

- Aim 1: Pilot-testing (find and correct problems in procedures)
- **Aim 2: Feasibility study (estimate a SD and its 95%CI)**
- Aim 3: Feasibility study (evaluate tolerability and retention)
- Aim 4: Small Exploratory Study (initial data for grant proposal, and generation / refinement of scientific hypotheses. )

# Pilot study ( $N_1$ ) to estimate SD for future study

---

## Problems with Small Pilot Studies

- If estimators in a study with a small sample size lacks precision, what happens if the study is used to estimate the SD in order to plan a future study based on the anticipated power of a test of interest ?

Let's find out ....

# Pilot study ( $N_1$ ) to estimate SD for future study

---

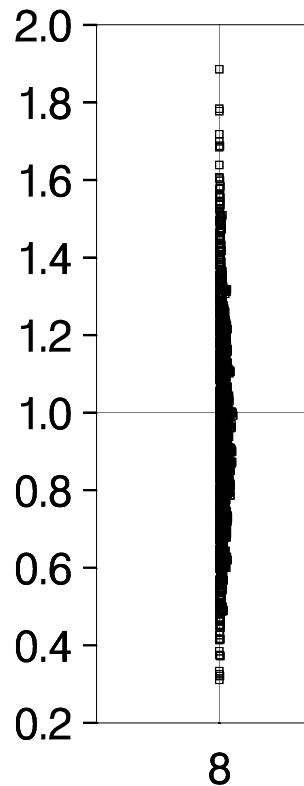
- $N_1$  is the sample size of the pilot study
- Pilot study ( $N_1$ ) yields  $\rightarrow$  SD and the 95%CI about the SD
- SD is the statistical estimate of the true std. dev. ( $\sigma$ ) in the target population
- Use SD estimates to compute  $\rightarrow N_2$  via a power calculation
- $N_2$  is the planned sample size of the future study

# Pilot study ( $N_1$ ) to estimate SD for future study

For a given value of  $N_1$ , 1000 pilot studies were simulated.

( Assume  $\sigma = 1$  without loss of generality )

**SD**



Example Simulation

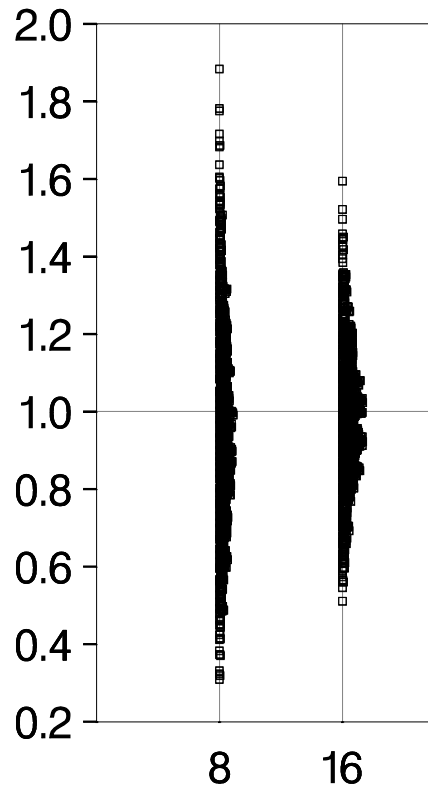
**$N_1$**

# Pilot study ( $N_1$ ) to estimate SD for future study

For a given value of  $N_1$ , 1000 pilot studies were simulated.

( Assume  $\sigma = 1$  without loss of generality )

**SD**



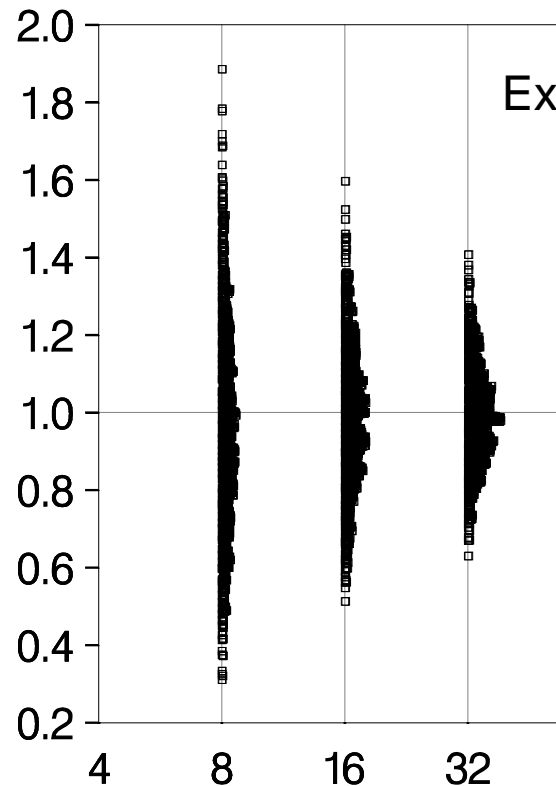
Example Simulation

# Pilot study ( $N_1$ ) to estimate SD for future study

For a given value of  $N_1$ , 1000 pilot studies were simulated.

( Assume  $\sigma = 1$  without loss of generality )

**SD**

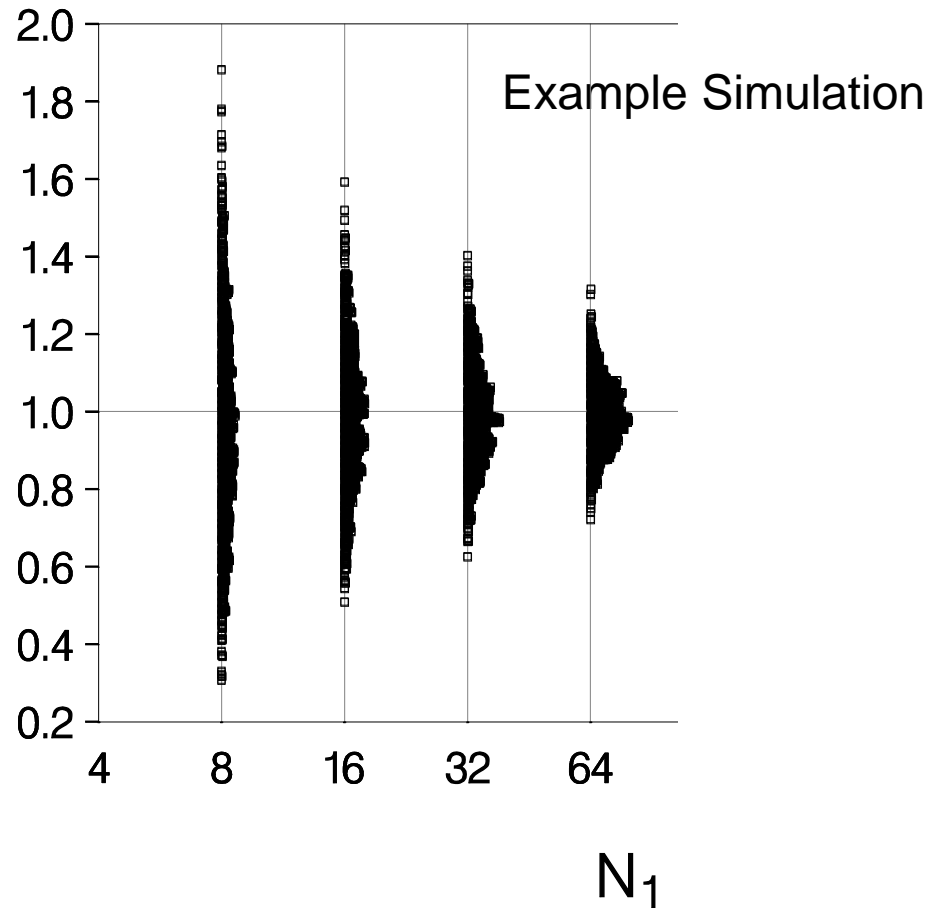


**$N_1$**

# Pilot study ( $N_1$ ) to estimate SD for future study

For a given value of  $N_1$ , 1000 pilot studies were simulated.  
( Assume  $\sigma = 1$  without loss of generality )

**SD**



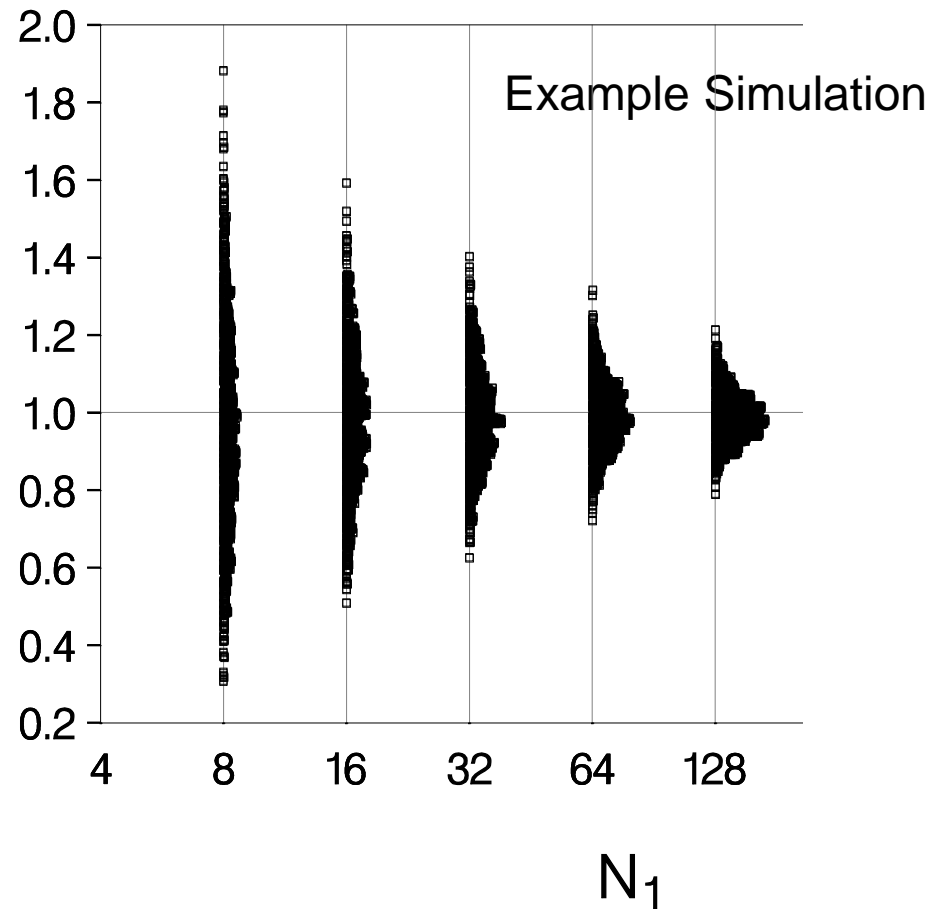


# Pilot study ( $N_1$ ) to estimate SD for future study

For a given value of  $N_1$ , 1000 pilot studies were simulated.

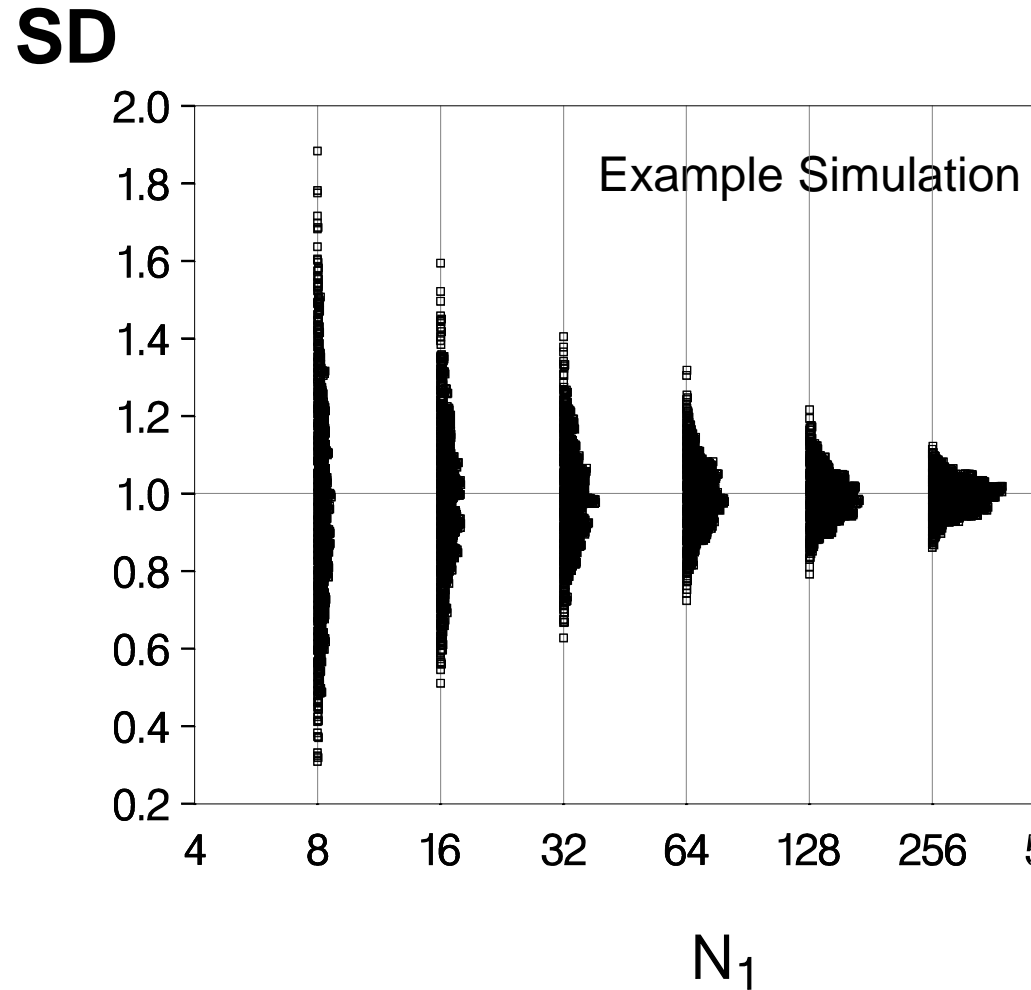
( Assume  $\sigma = 1$  without loss of generality )

**SD**



# Pilot study ( $N_1$ ) to estimate SD for future study

For a given value of  $N_1$ , 1000 pilot studies were simulated.  
( Assume  $\sigma = 1$  without loss of generality )



# Pilot study ( $N_1$ ) to estimate SD for future study

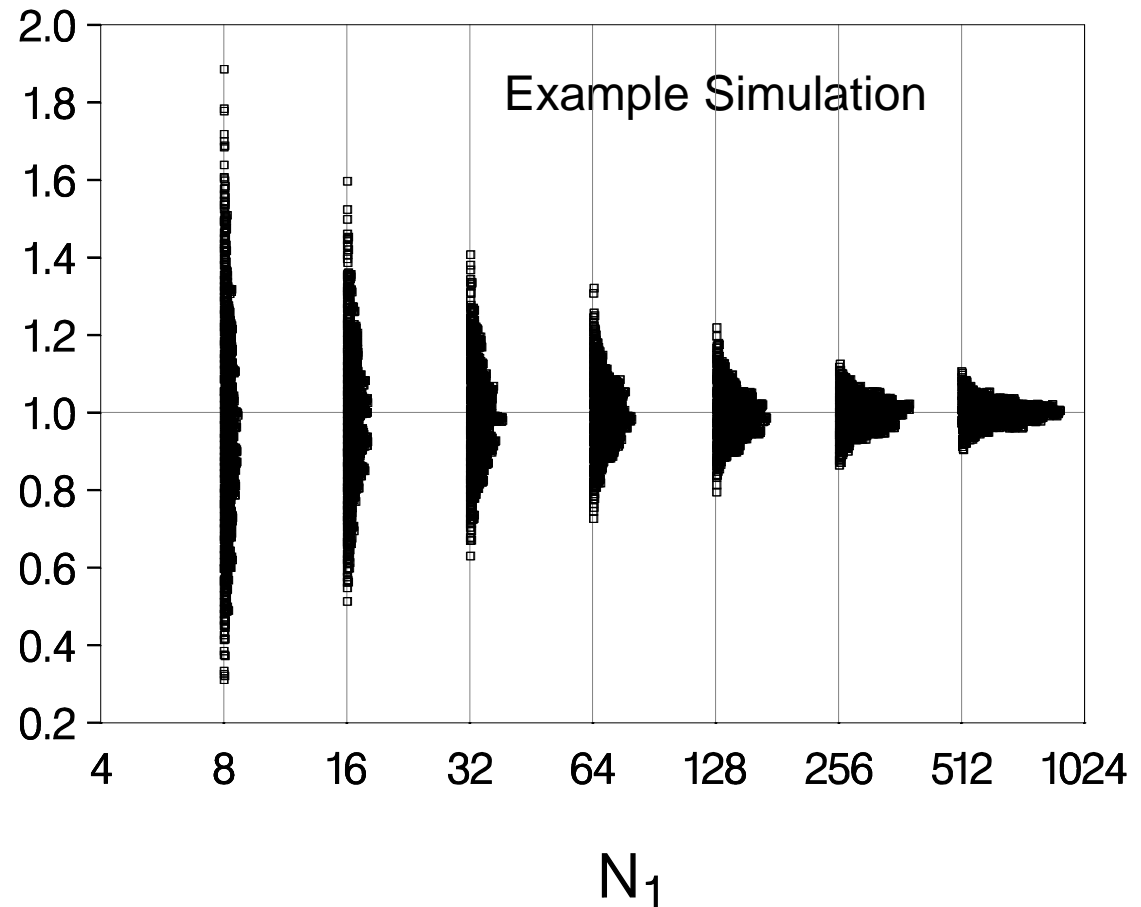
For a given value of  $N_1$ , 1000 pilot studies were simulated.

( Assume  $\sigma = 1$  without loss of generality )

Note:

$N_1$  on  $\log_2$  scale.

**SD**



# Pilot study ( $N_1$ ) to estimate SD for future study

- Pilot study sample size is  $N_1$
- Pilot study ( $N_1$ )  $\rightarrow$  estimate (SD) of population std. dev. ( $\sigma$ )
- Assumptions for the power calculation
  - ◆  $\sigma = \text{SD}$
  - ◆ t-test of  $H_0: \Delta = 0$
  - ◆  $\Delta = \Delta_0$  in the target population
  - ◆ size of t-test is  $\alpha = 0.05$
  - ◆ desire 90% power
- SD  $\rightarrow N_2$  which is the target sample size of the future study
- $\Pr[ \text{p-value} < \alpha \mid N_2, \Delta_0, \sigma = \text{SD} ] = 90\%$

# Pilot study ( $N_1$ ) to estimate SD for future study

Assume  $\sigma = \text{SD}$

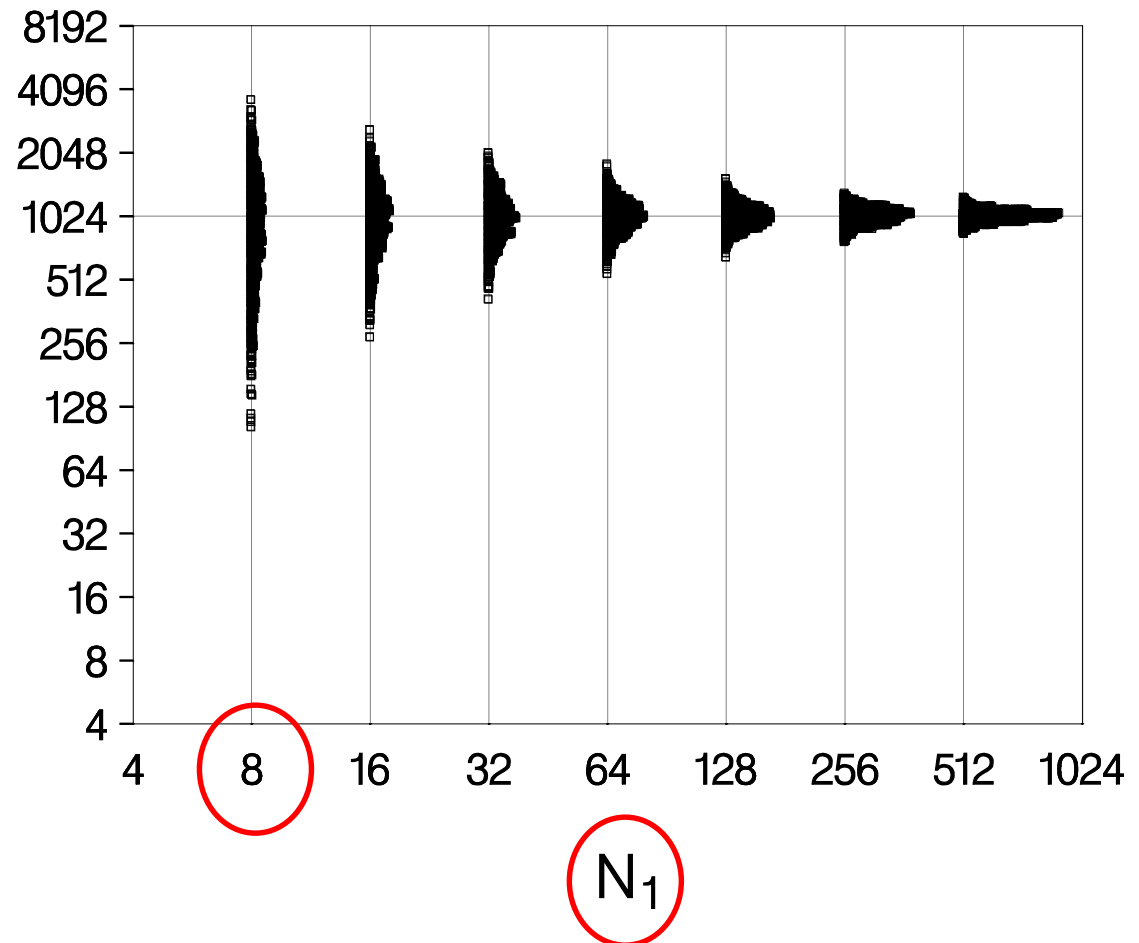
$N_2$

Target = 1024

Criterion: Power=90%	
$\Delta_0$	Target $N_2$
<b>0.20285</b>	<b>1024</b>

When  $\Delta_0 = 0.20285$  (small !) we need  $N_2 = 1024$  (large !) to achieve power=90%.

But when  $N_1 = 8$  the SD estimate is unreliable and the power calculation may suggest  $N_2$  values as large as 4000 or as small as 100 !



# Pilot study ( $N_1$ ) to estimate SD for future study

Assume  $\sigma = \text{SD}$

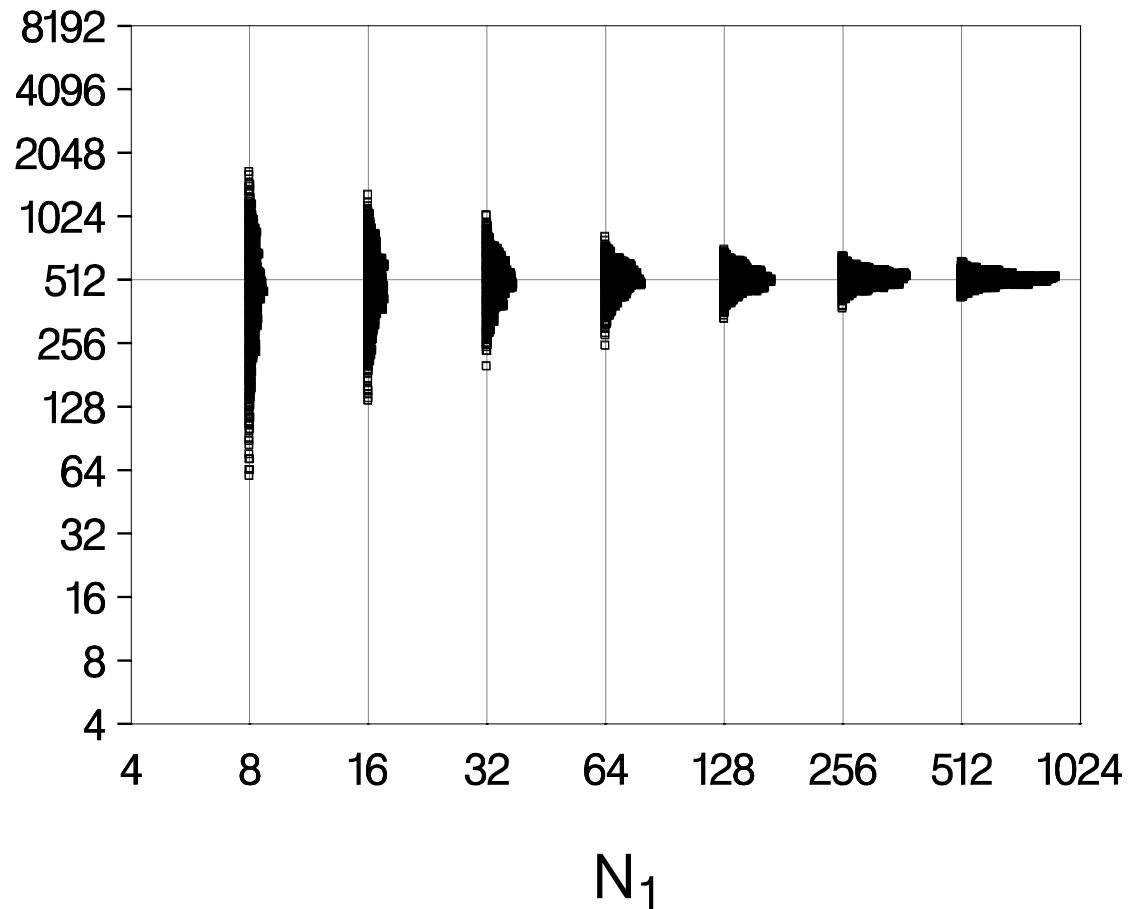
$N_2$

Target = 512

Criterion: Power=90%	
$\Delta_0$	Target $N_2$
0.20285	1024
<b>0.28750</b>	<b>512</b>

When  $\Delta_0 = 0.28750$   
we need  $N_2 = 512$   
to achieve power=90%.

The SD estimate is not very  
reliable even for  $N_1 = 64$



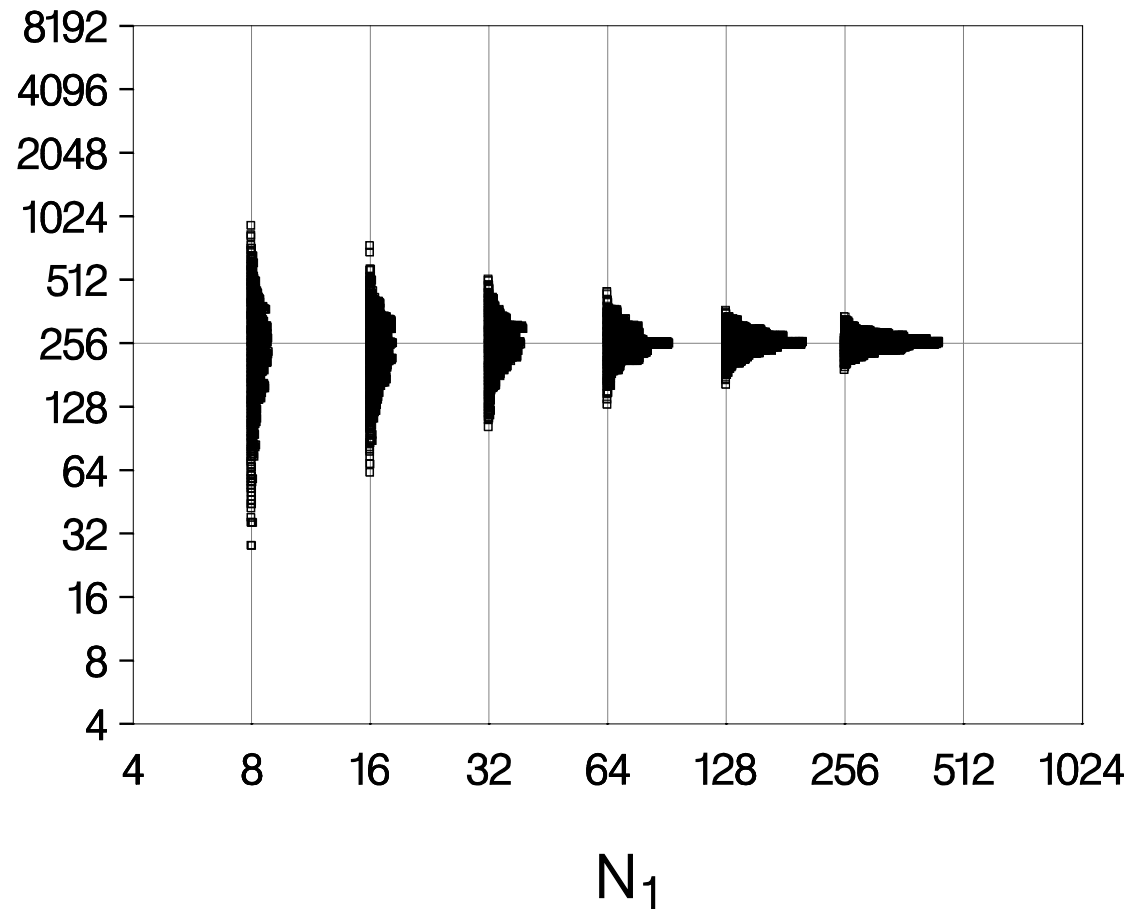
# Pilot study ( $N_1$ ) to estimate SD for future study

Assume  $\sigma = \text{SD}$

$N_2$

Target = 256

Criterion: Power=90%	
$\Delta_o$	Target $N_2$
0.20285	1024
0.28750	512
<b>0.40720</b>	<b>256</b>



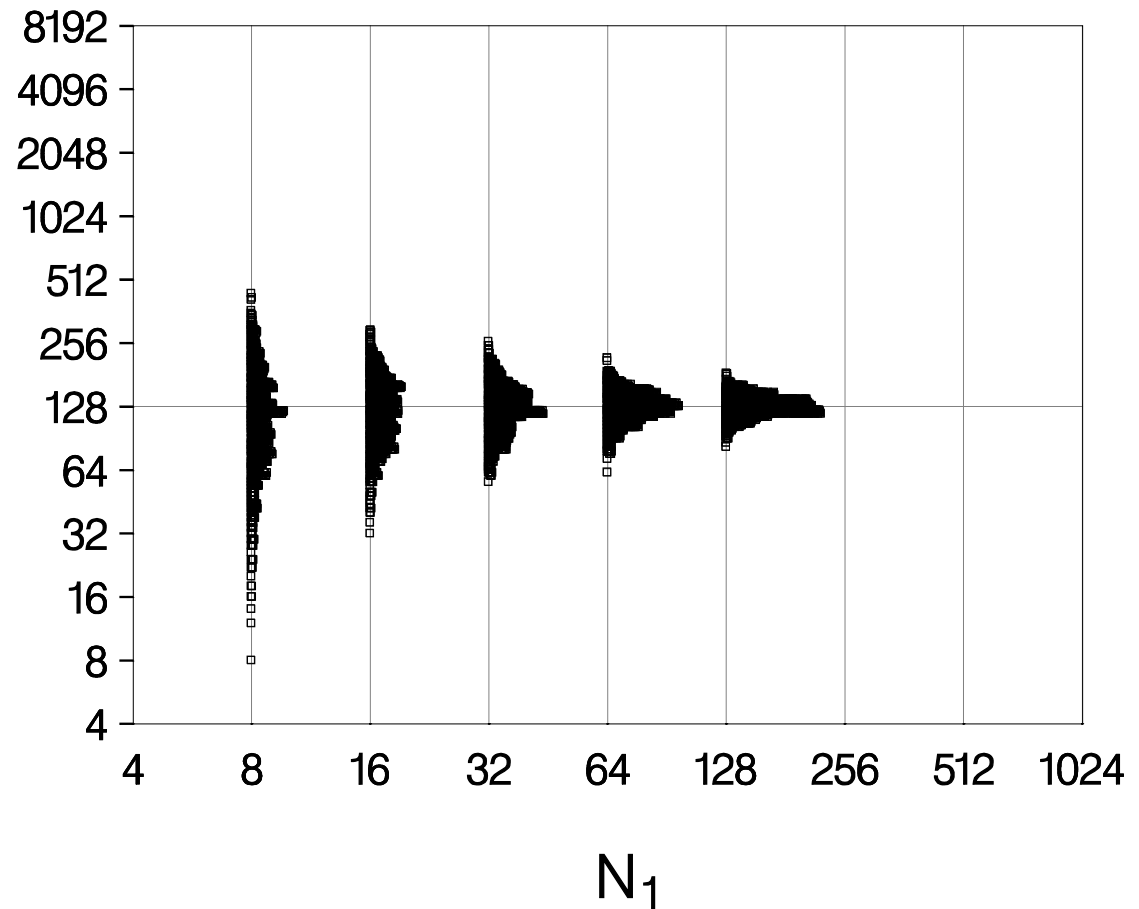
# Pilot study ( $N_1$ ) to estimate SD for future study

Assume  $\sigma = \text{SD}$

$N_2$

Target = 128

Criterion: Power=90%	
$\Delta_0$	Target $N_2$
0.20285	1024
0.28750	512
0.40720	256
<b>0.58000</b>	<b>128</b>





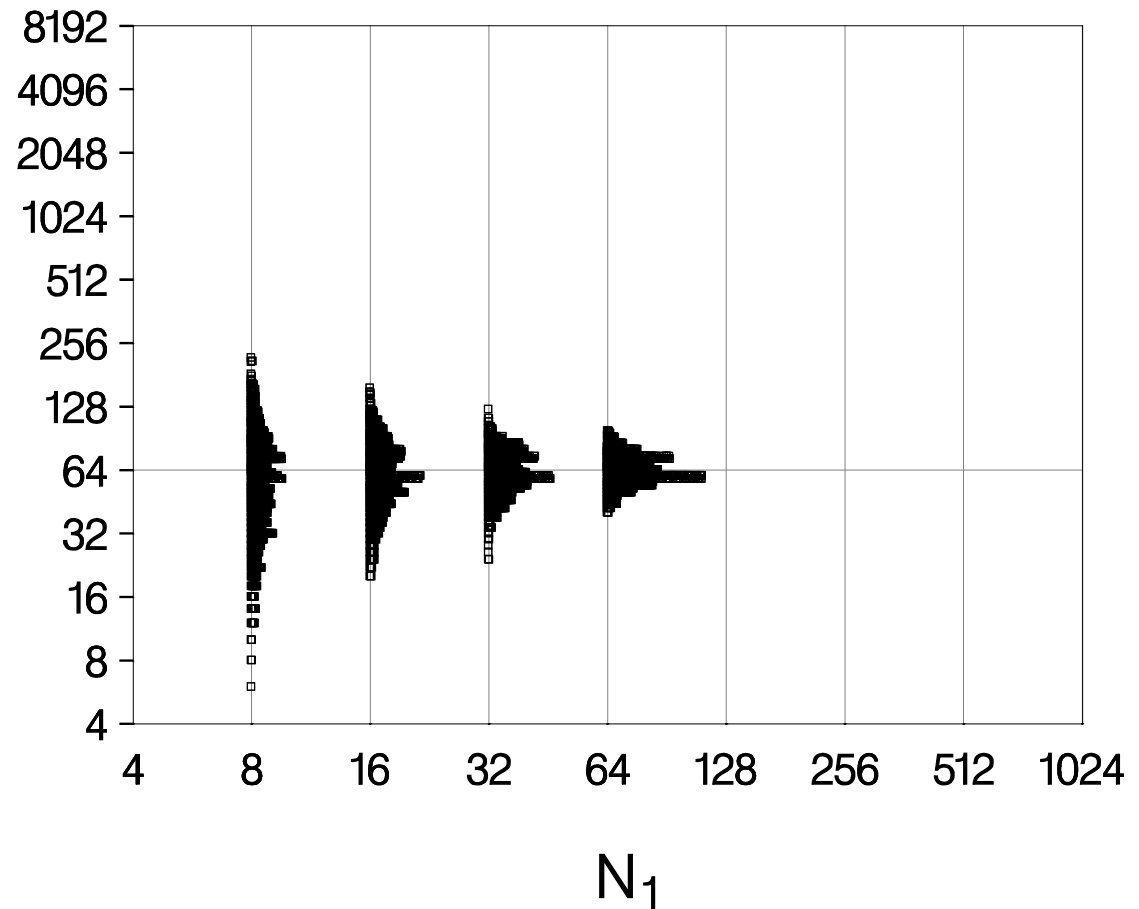
# Pilot study ( $N_1$ ) to estimate SD for future study

Assume  $\sigma = \text{SD}$

$N_2$

Target = 64

Criterion: Power=90%	
$\Delta_o$	Target $N_2$
0.20285	1024
0.28750	512
0.40720	256
0.58000	128
<b>0.83000</b>	<b>64</b>



When  $\Delta_o = 0.83000$   
we only need  $N_2 = 64$   
to achieve power=90%.

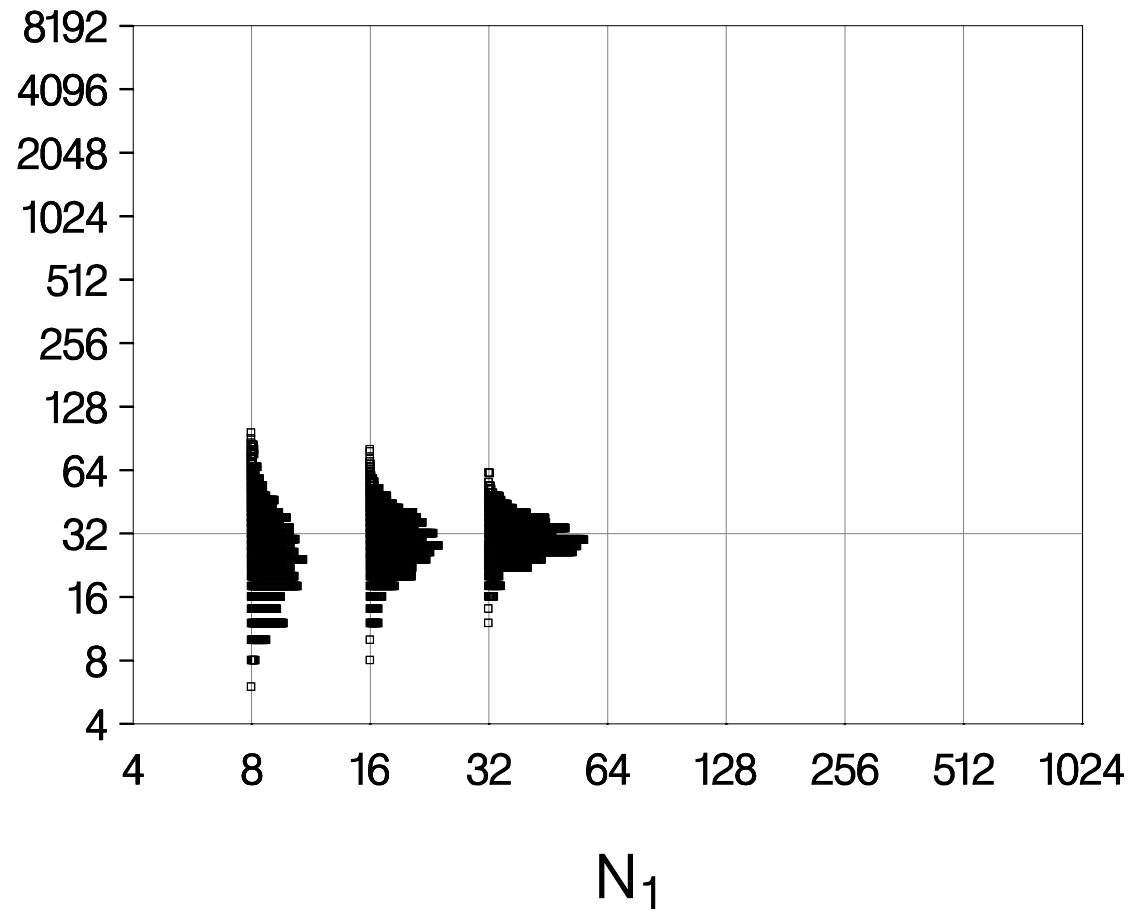
# Pilot study ( $N_1$ ) to estimate SD for future study

Assume  $\sigma = \text{SD}$

$N_2$

Target = 32

Criterion: Power=90%	
$\Delta_0$	Target $N_2$
0.20285	1024
0.28750	512
0.40720	256
0.58000	128
0.83000	64
<b>1.20000</b>	<b>32</b>



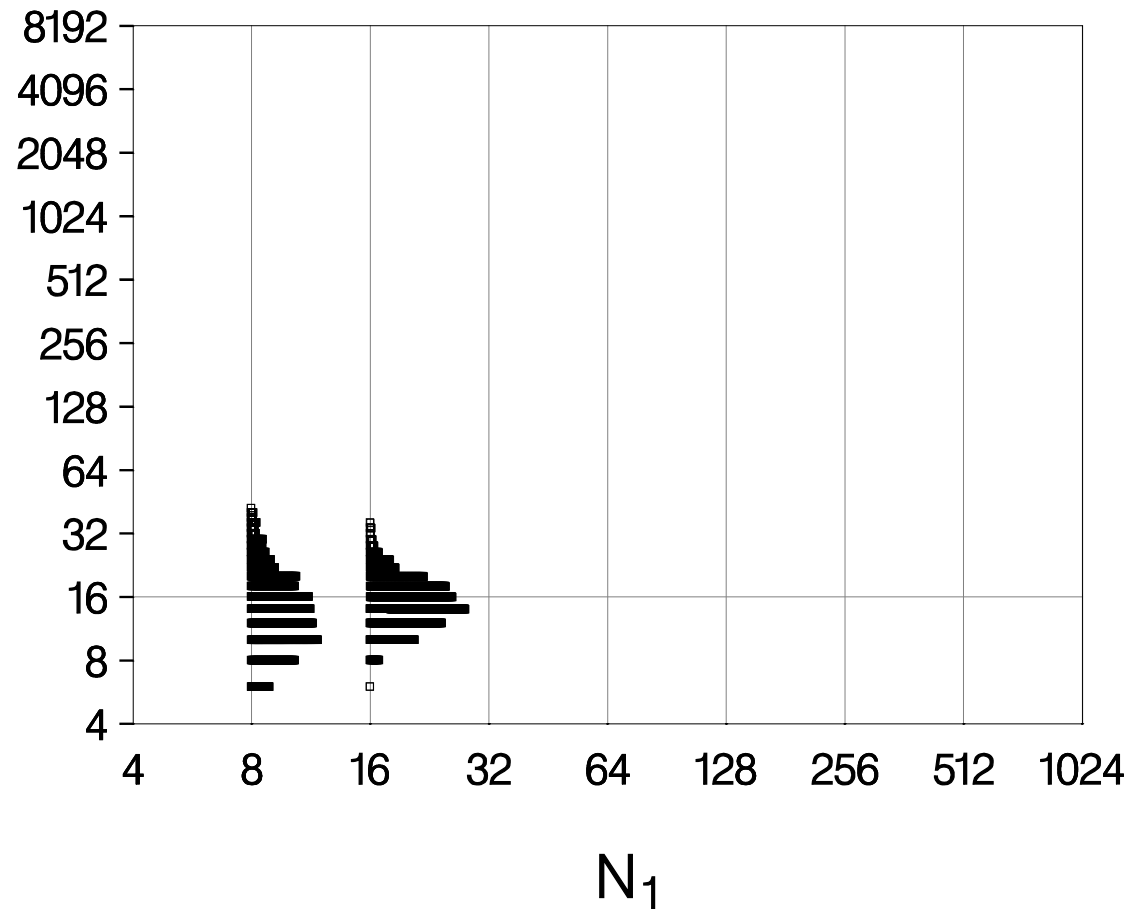
# Pilot study ( $N_1$ ) to estimate SD for future study

Assume  $\sigma = \text{SD}$

$N_2$

Target = 16

Criterion: Power=90%	
$\Delta_o$	Target $N_2$
0.20285	1024
0.28750	512
0.40720	256
0.58000	128
0.83000	64
1.20000	32
<b>1.80000</b>	<b>16</b>



# Pilot study ( $N_1$ ) $\rightarrow$ SD $\rightarrow$ $N_2$ $\rightarrow$ anticipated power

Assume  $\sigma = \text{SD}$

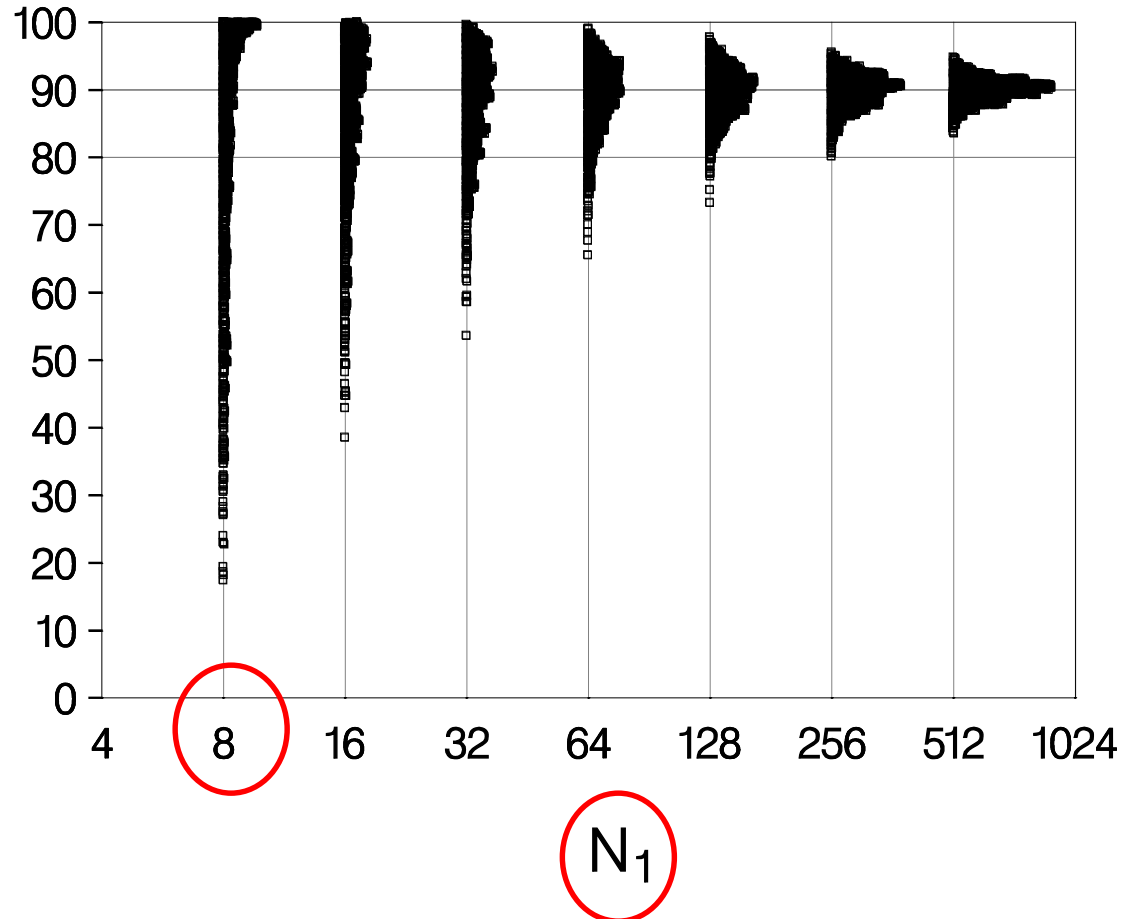
Criterion: Power=90%	
$\Delta_0$	Target $N_2$
<b>0.20285</b>	<b>1024</b>

When  $\Delta_0 = 0.20285$  (small !)  
we need  $N_2 = 1024$  (large !)  
to achieve power=90%.

But when  $N_1 = 8$  the SD  
estimate is unreliable and the  
resulting  $N_2$  values can be so  
far off-target that the actual  
power level is not near 90% !

Power

Target = 1024



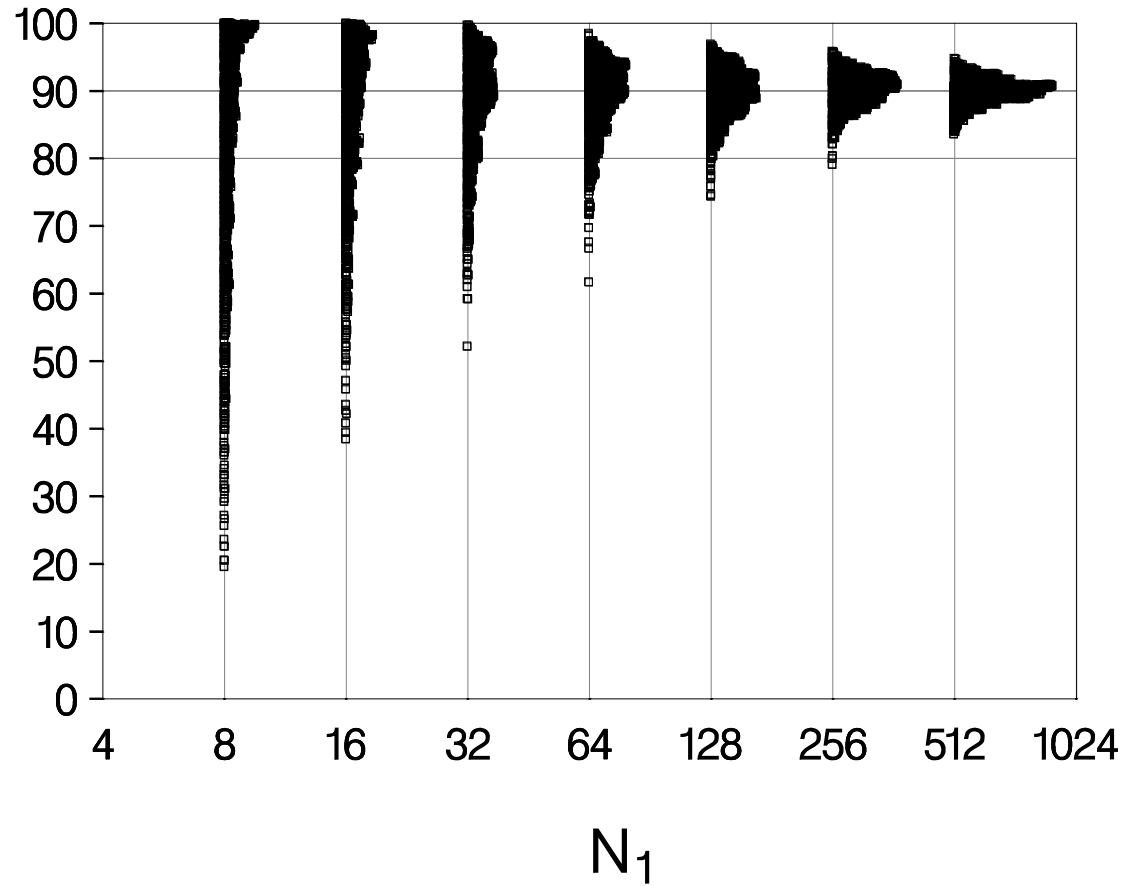
# Pilot study ( $N_1$ ) $\rightarrow$ SD $\rightarrow$ $N_2$ $\rightarrow$ anticipated power

Assume  $\sigma = \text{SD}$

Power

Target = 512

Criterion: Power=90%	
$\Delta_o$	Target $N_2$
0.20285	1024
<b>0.28750</b>	<b>512</b>
0.40720	256
0.58000	128
0.83000	64
1.20000	32
1.80000	16



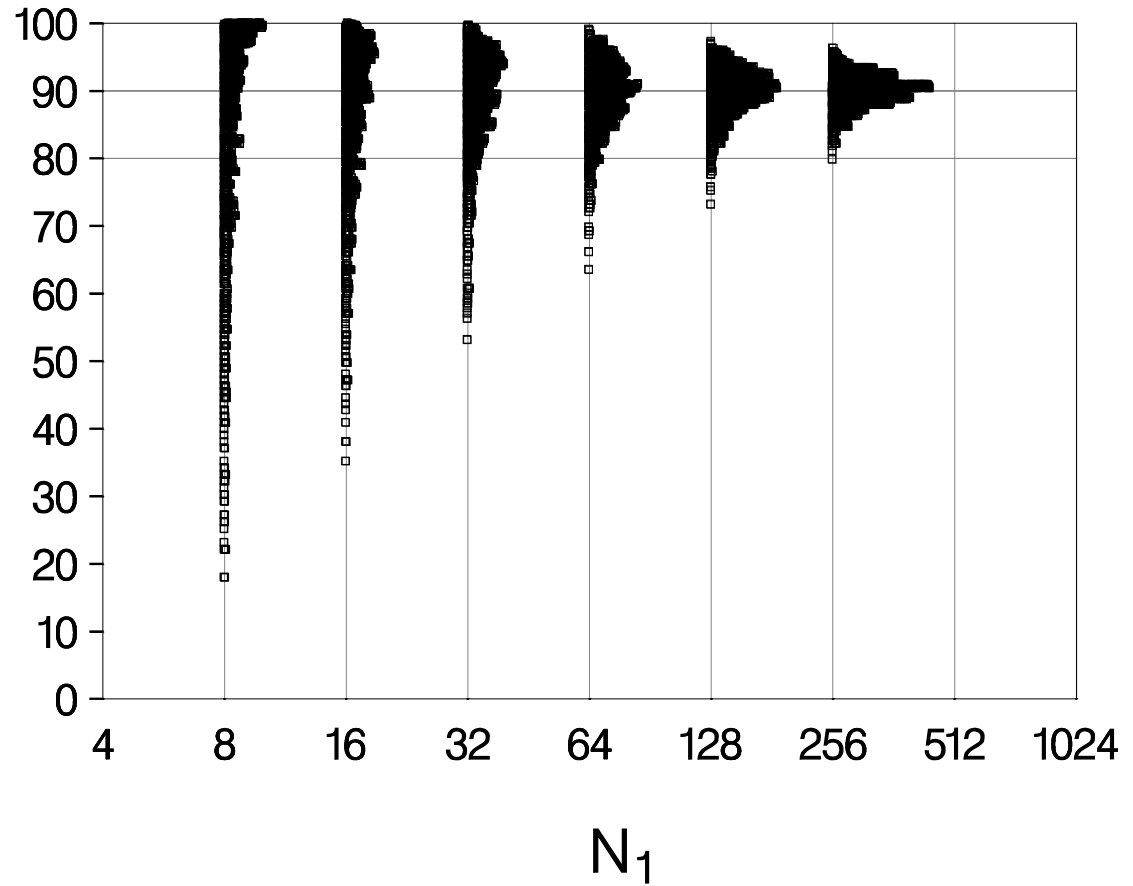
# Pilot study ( $N_1$ ) $\rightarrow$ SD $\rightarrow$ $N_2$ $\rightarrow$ anticipated power

Assume  $\sigma = \text{SD}$

Power

Target = 256

Criterion: Power=90%	
$\Delta_o$	Target $N_2$
0.20285	1024
0.28750	512
<b>0.40720</b>	<b>256</b>
0.58000	128
0.83000	64
1.20000	32
1.80000	16



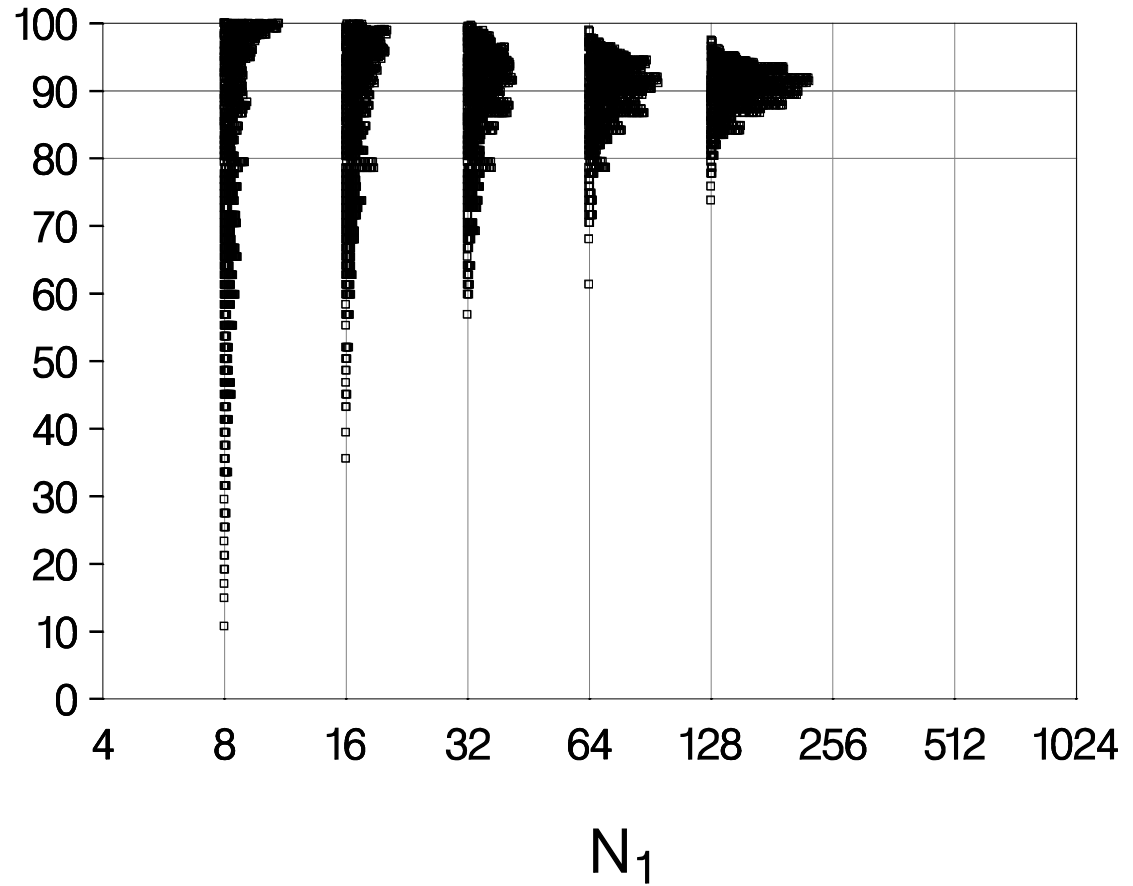
# Pilot study ( $N_1$ ) $\rightarrow$ SD $\rightarrow$ $N_2$ $\rightarrow$ anticipated power

Assume  $\sigma = \text{SD}$

Power

Target = 128

Criterion: Power=90%	
$\Delta_o$	Target $N_2$
0.20285	1024
0.28750	512
0.40720	256
<b>0.58000</b>	<b>128</b>
0.83000	64
1.20000	32
1.80000	16



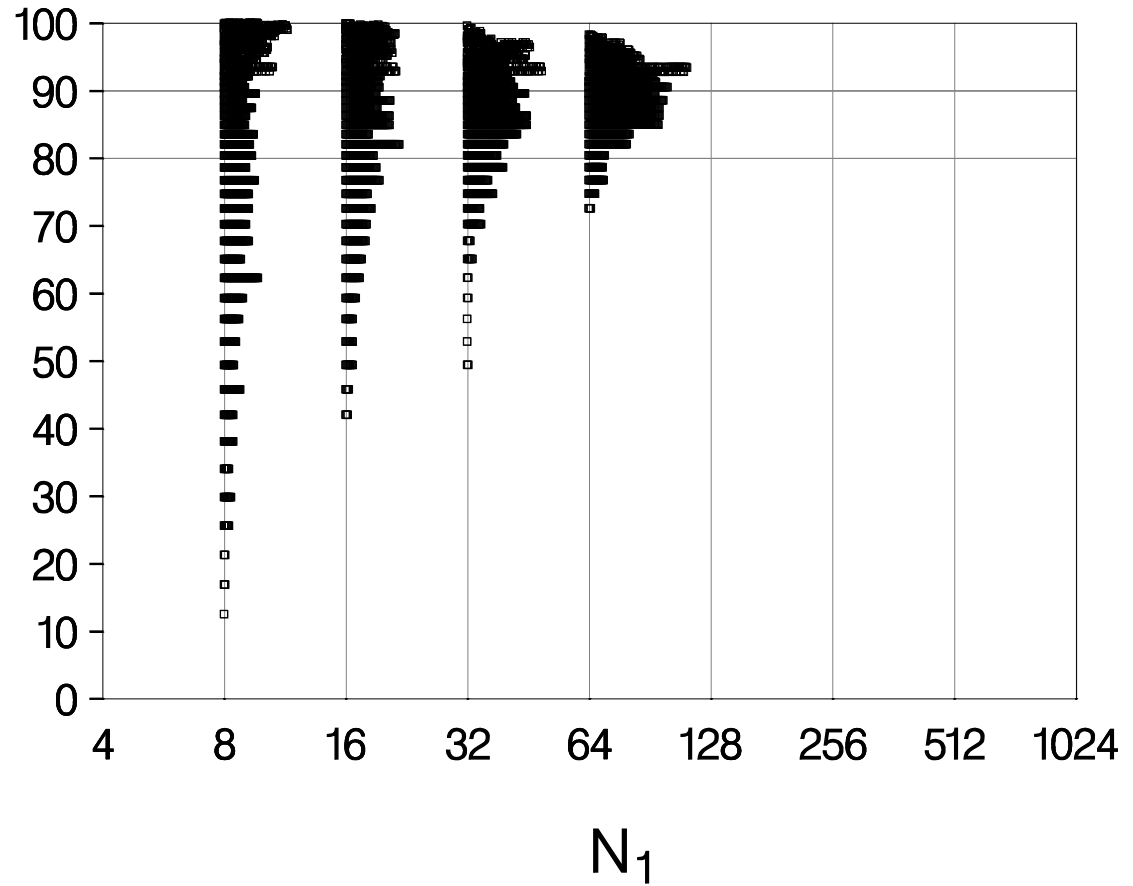
# Pilot study ( $N_1$ ) $\rightarrow$ SD $\rightarrow$ $N_2$ $\rightarrow$ anticipated power

Assume  $\sigma = \text{SD}$

Power

Target = 64

Criterion: Power=90%	
$\Delta_o$	Target $N_2$
0.20285	1024
0.28750	512
0.40720	256
0.58000	128
<b>0.83000</b>	<b>64</b>
1.20000	32
1.80000	16





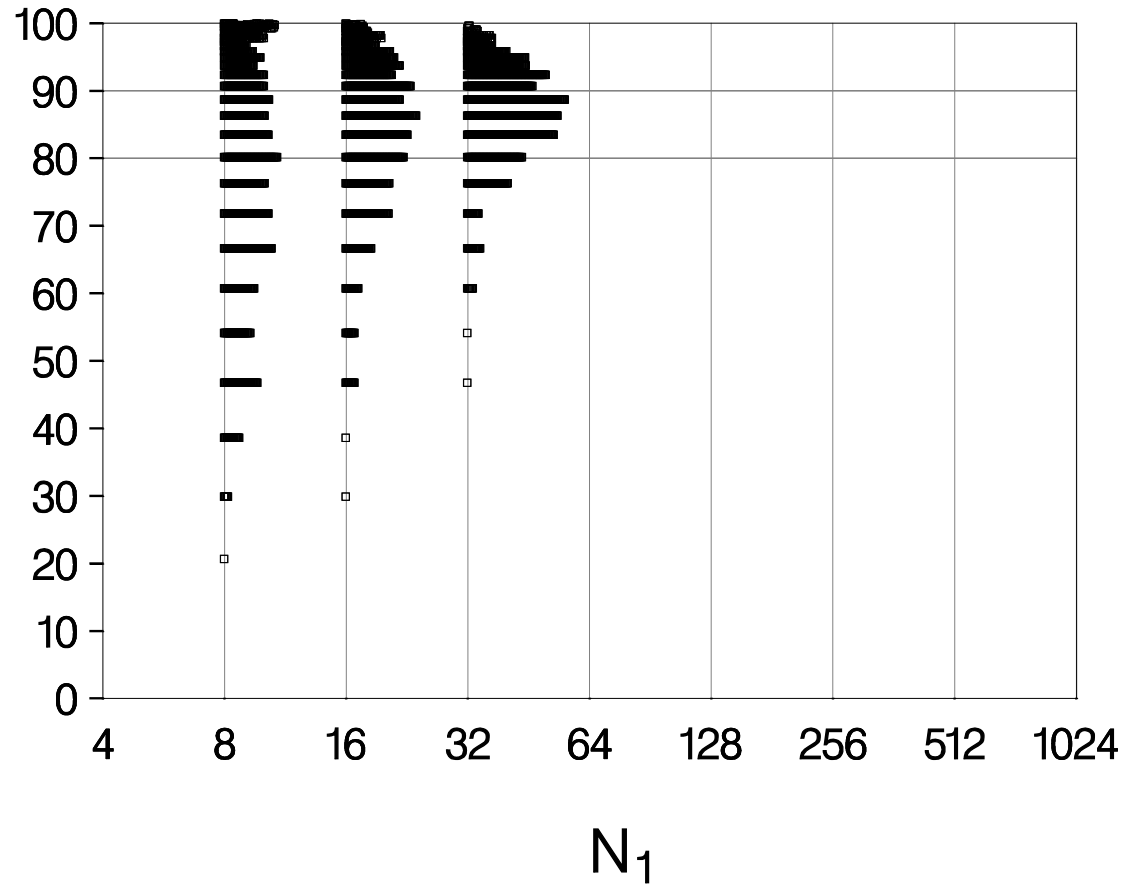
# Pilot study ( $N_1$ ) $\rightarrow$ SD $\rightarrow$ $N_2$ $\rightarrow$ anticipated power

Assume  $\sigma = \text{SD}$

Power

Target = 32

Criterion: Power=90%	
$\Delta_o$	Target $N_2$
0.20285	1024
0.28750	512
0.40720	256
0.58000	128
0.83000	64
<b>1.20000</b>	<b>32</b>
1.80000	16



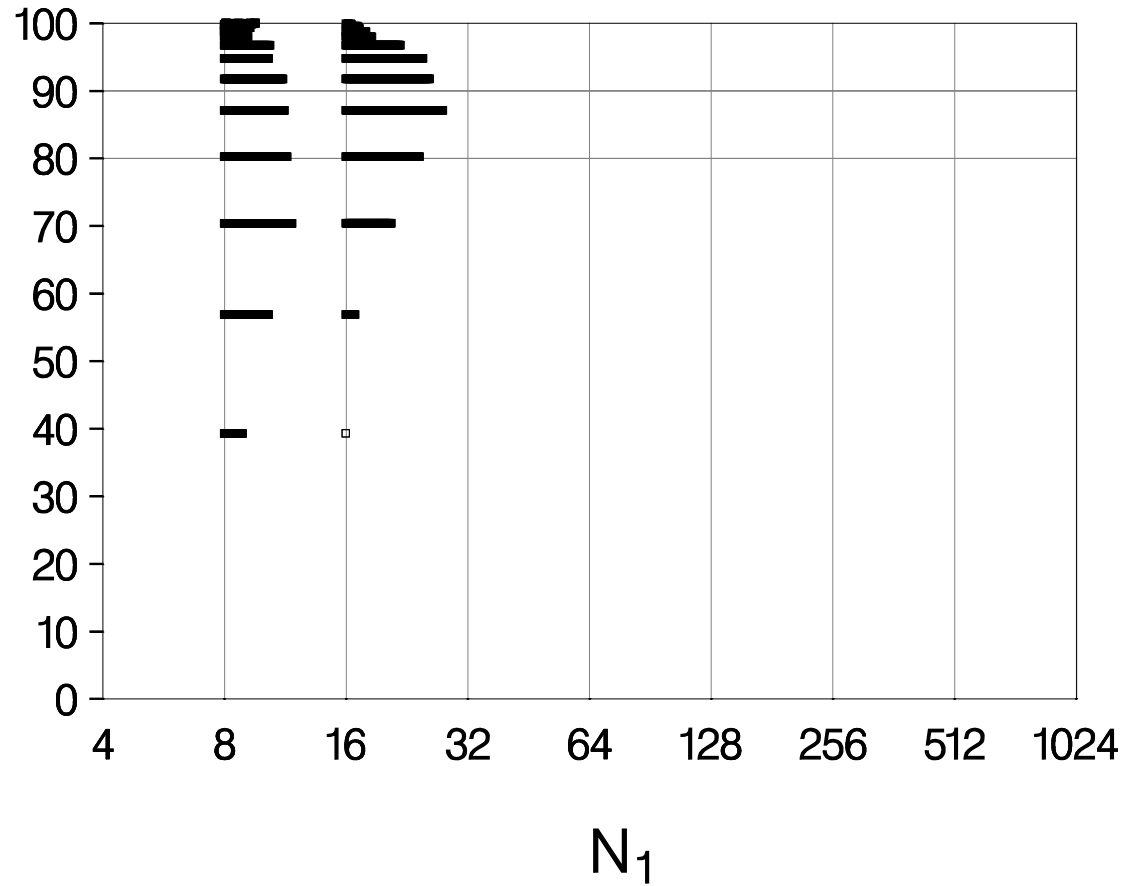
# Pilot study ( $N_1$ ) $\rightarrow$ SD $\rightarrow$ $N_2$ $\rightarrow$ anticipated power

Assume  $\sigma = \text{SD}$

Power

Target = 16

Criterion: Power=90%	
$\Delta_o$	Target $N_2$
0.20285	1024
0.28750	512
0.40720	256
0.58000	128
0.83000	64
1.20000	32
<b>1.80000</b>	<b>16</b>



# Pilot study ( $N_1$ ) $\rightarrow$ SD $\rightarrow$ $N_2$ $\rightarrow$ anticipated power

---

## The results illustrate a problem

- The task of estimating the optimal sample size (**Target- $N_2$** ) is more difficult than might be expected.
- Small external pilot studies in particular can suggest  $N_2$  values that are far from the target.
- Even the use of a large preliminary study is subject to a substantial likelihood of choosing  $N_2$  far from the target.

# Pilot study ( $N_1$ ) $\rightarrow$ SD $\rightarrow$ $N_2$ $\rightarrow$ anticipated power

---

## Question

- In practice, if large  $N_2$  values are deemed infeasible, but small values of  $N_2$  are deemed easily feasible, what will happen?

# Pilot study ( $N_1$ ) $\rightarrow$ SD $\rightarrow$ $N_2$ $\rightarrow$ anticipated power

## Question

- In practice, if large  $N_2$  values are less feasible, but small values of  $N_2$  are deemed easily feasible, what will happen?

- The future study will go forward only if  $N_2$  is small.
- The result can be an excess of inconclusive and uninformative studies.

# Pilot study ( $N_1$ ) $\rightarrow$ SD $\rightarrow$ $N_2$ $\rightarrow$ anticipated power

---

## To avoid the problem...

- Make use of information (e.g., about SD) in previously published studies
  - ◆ Numerous studies of the outcome variable of interest are highly informative even if they were not studies of the novel treatment regimen and subpopulation now of interest to you.
- Consider use of
  - ◆ internal-pilot study designs,
  - ◆ group-sequential study designs,
  - ◆ other kinds of adaptive study designs.
- Give serious attention to (perhaps large) uncertainty indicated by the C.I.s for inputs (e.g., SD) and C.I.s for estimates of power and the margin of error. When those C.I.s are very wide, understand that the sample size analysis is highly uncertain.

# Essential concepts, Best practices, Pitfalls, Speedy IRB approval

