Sample Size Considerations

Todd Alonzo, PhD

1

1

Acknowledgements

Thanks to Nancy Obuchowski for previous version of this presentation.

2

2

Outline

- 4 steps in process to determining sample size
- Two example studies
  - Estimate sensitivity
  - Compare two modalities' ROC areas

3

3

Why do Sample Size Calculations?

1. To determine if a study is feasible
2. To plan for your study's needs
3. To minimize the risk of making the wrong conclusion from your study

4

4

State
statistical hypothesis
and/or
what you want to
estimate

The process of sample size calculations takes place after you have specified your study's primary objective.

5

5

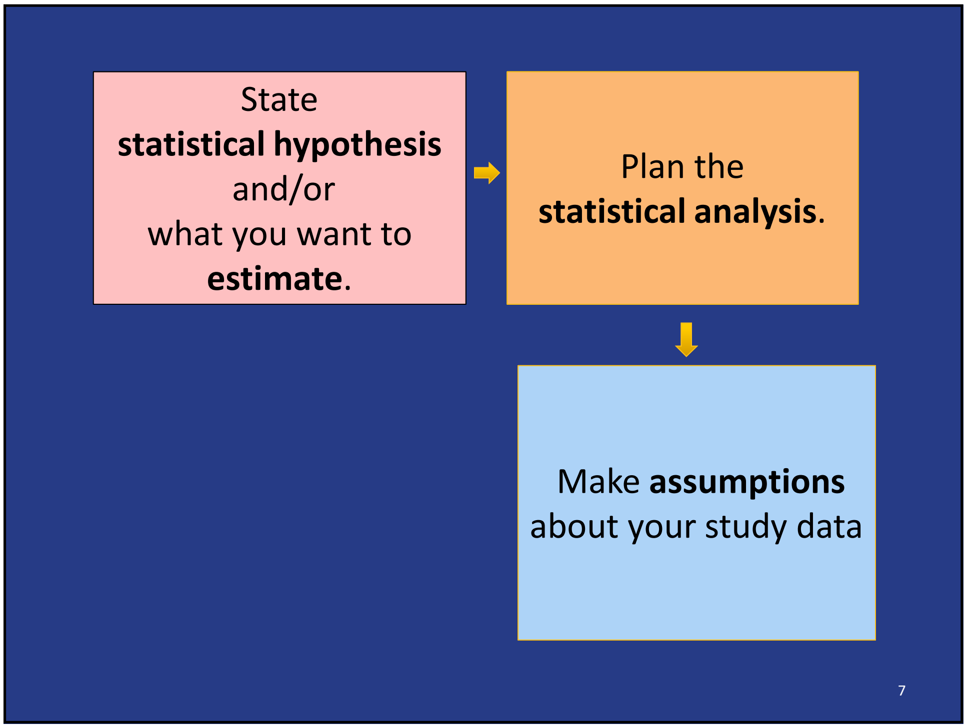
State
statistical hypothesis
and/or
what you want to
estimate

➡

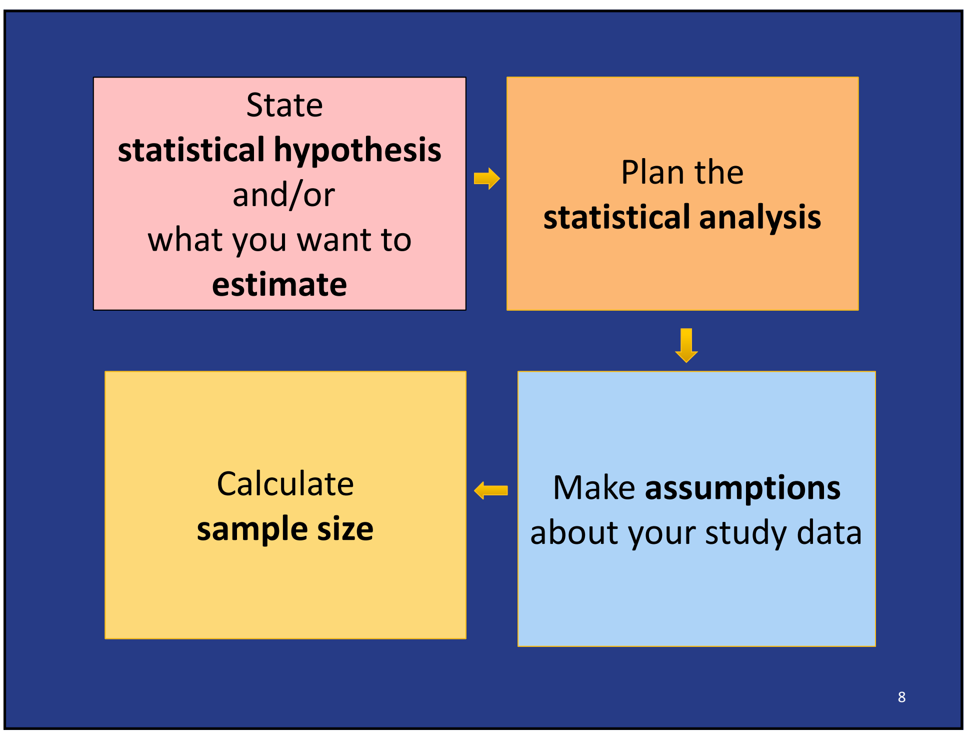
Consider the **study design** and plan for **statistical analysis**

6

6


State
statistical hypothesis
and/or
what you want to
estimate.
Plan the
statistical analysis.
Make assumptions
about your study data
7

7


State
statistical hypothesis
and/or
what you want to
estimate
Plan the
statistical analysis
Make assumptions
about your study data
Calculate
sample size
8

8

Example 1: Study to estimate sensitivity of new modality

- One reader will score each patient using 5-point scale
- Define positive test result as score $\geq$3
- Retrospective design

9

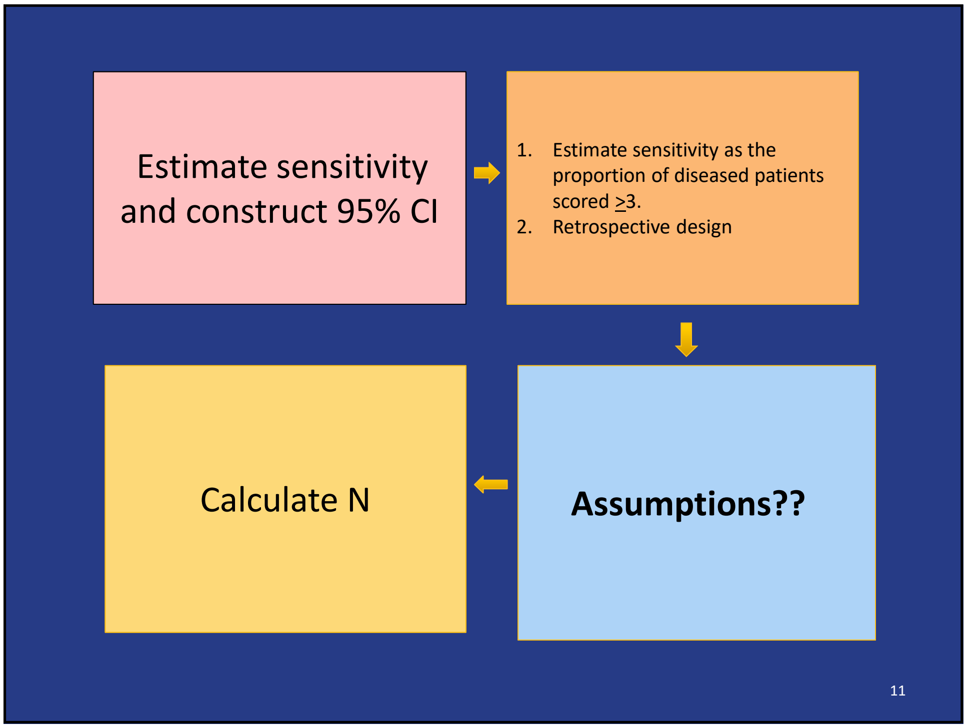
9

Example 1 – primary objective

"To estimate and compare the breast-level sensitivity and specificity of board-certified mammographers interpreting breast MRI vs. mammograms of high-risk women."

10

10



11

Conversation with statistician

Radiologist:

"I want to estimate sensitivity of my new modality"

Statistician:

"How sensitive do you expect it to be?"

Radiologist:

"Maybe between 0.7 and 0.9"

Statistician:

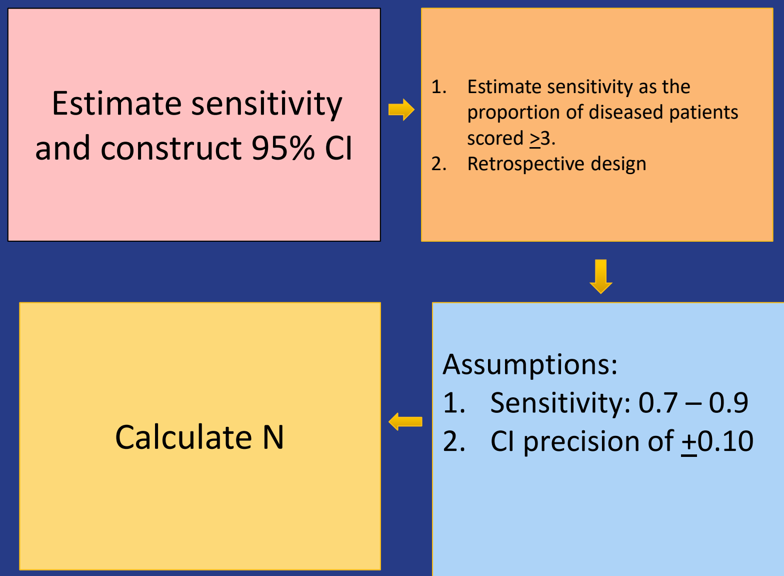
"Ok. We should construct a CI. How precisely do you want to estimate sensitivity?"

Radiologist:

"To within 0.10 would be ok."

12

12



13

13

Sample Size Formula for Constructing CI

$$N = [(z_{\alpha/2})^2 \ (Var)] / (L)^2$$

$z_{\alpha/2}$ = z-value to specify confidence level (usually 95% confidence)

Var = variance

L = Half-width of CI

14

14

Some Useful z-values

For 95% confidence level	$z_{\alpha/2} = 1.96$
For 5% type I error rate	$z_{\alpha/2} = 1.96$
For 80% power	$z_{\beta} = 0.84$
For 90% power	$z_{\beta} = 1.28$

15

15

Example 1

Variability for a proportion: $p \times (1-p)$

p	Variability
0.5	0.25
0.6	0.24
0.7	0.21
0.8	0.16
0.9	0.09

16

16

Example 1

$$N = [(1.96)^2 (0.7 \times 0.3)] / (0.10)^2 = 80.7$$

So we need 81 patients with disease.

17

17

Impact of Study Design on N

Retrospective Study: Identify # diseased subjects needed for study: $N_{DIS} = 81$. Identify these diseased subjects based on reference standard.

18

18

Impact of Study Design on N

Retrospective Study: Identify # diseased subjects needed for study: N_{DIS} = 81. Identify these diseased subjects based on reference standard.

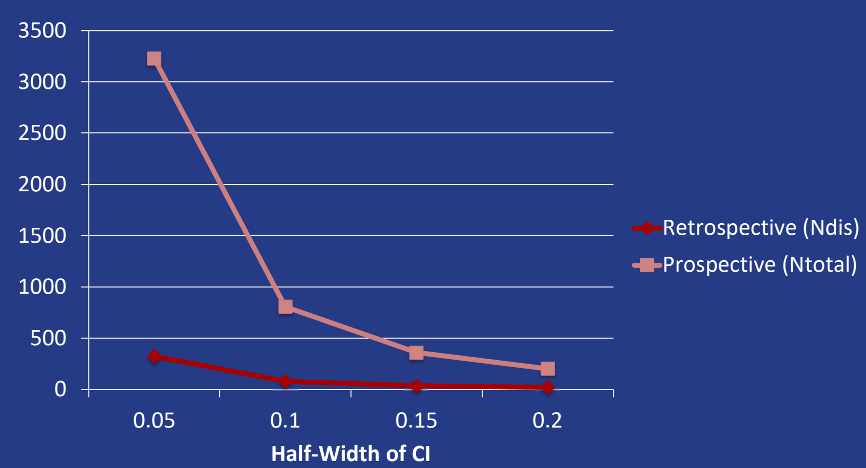
Prospective Study: Must consider prevalence rate in population. Suppose prevalence rate is 10%.

total N = N_{DIS} / Prevalence = 80.7/0.10 = 807

19

19

Number of Subjects Required



20

20

Example 2: Study to Compare ROC Areas of Two Modalities

- Suppose you want to compare a new modality with a standard modality
- Readers will assign a confidence score (0-100)
- It's an accuracy study, so confidence scores will be compared against reference standard
- Retrospective design

21

21

Alternative Hypotheses

Superiority	One modality is better than the other
Non-Inferiority	New modality is at least as good as the standard modality
Equivalence	New modality has same performance as the standard modality

22

22

Statistical Hypotheses

Superiority	H_o: $AUC_{new} = AUC_{old}$ H_A: $AUC_{new} \neq AUC_{old}$
Non-Inferiority	
Equivalence	

23

23

Statistical Hypotheses

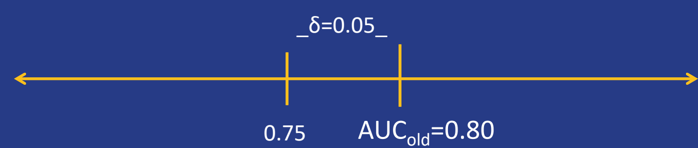
Superiority	H_o: $AUC_{new} = AUC_{old}$ H_A: $AUC_{new} \neq AUC_{old}$
Non-Inferiority	H_o: $AUC_{new} + \delta \leq AUC_{old}$ H_A: $AUC_{new} + \delta > AUC_{old}$
Equivalence	

24

24

Non-inferiority criterion

If the new test is easier to use than the old test (e.g. fewer complications, less expensive, less radiation, less time, less personnel), it doesn't have to be quite so accurate:



25

25

Statistical Hypotheses

Superiority	
Non-Inferiority	
Equivalence	H_o: $(AUC_{new}-AUC_{old}) \leq \delta_L$ or $(AUC_{new}-AUC_{old}) \geq \delta_U$ H_A: $\delta_L < (AUC_{new}-AUC_{old}) < \delta_U$

26

26

Equivalence criteria

The difference between the new and old test must be between the lower and the upper equivalence bounds:

27

27

H_o: $AUC_{new} + \delta \leq AUC_{old}$

H_A: $AUC_{new} + \delta > AUC_{old}$

1. Nonparametric estimates
2. Non-inferiority test
3. 5% type I error rate
4. Retrospective design
5. Equal # diseased & nondiseased subjects

Assumptions??

Calculate **sample size for 1-reader study**

28

28

Conversation with statistician

Radiologist:

"I want to test if my new modality is non-inferior to the old modality."

"Maybe about AUC=0.8."

"I think it is about the same."

"Yes. A paired subject design."

Statistician:

"How accurate is the old test?"

"How accurate is the new test?"

"Will the same subjects undergo the new and old test?"

"Ok. For sample size calculation, I'll assume moderate correlation between the 2 tests."

29

29

Accuracy of New Test

If $AUC_{new} >>> AUC_{old}$ ➔ We don't need a large N

.

.

.

If $AUC_{new} < AUC_{old}$ ➔ We need a much larger N

30

30

H_o: $AUC_{new} + \delta \leq AUC_{old}$

H_A: $AUC_{new} + \delta > AUC_{old}$

1. Nonparametric estimates
2. Non-inferiority test
3. 5% type I error rate
4. Retrospective; PAIRED
5. Equal # diseased & nondiseased subject

1. Old test average AUC = 0.8
2. New test average AUC =0.8
3. δ=0.05
4. $r_{subjects}$ =0.5

Calculate
sample size for 1-reader study

31

31

Sample Size Formula for One-Reader Study

$$N_{dis} = \frac{(z_{\alpha/2} + z_{\beta})^2 \times [\ 2 \times \{Var_{subject} \times (1-r_{subject})\ \}\ ]}{(\theta_{new} - \theta_{old} + \delta)^2}$$

32

32

Conclusions

Truth	Fail to reject Null	Reject Null
Null is truth (inferior)	Correct conclusion	Incorrect conclusion
Alternative is truth (non-inferior)	Incorrect conclusion	Correct conclusion

33

33

Type I and II errors

Truth	Fail to reject Null	Reject Null
Null is truth	Correct conclusion	Type I error
Alternative is truth	Type II error	POWER

34

34

Some Useful z-values

For 95% confidence level	$z_{\alpha/2}$ = 1.96
For 5% type I error rate	$z_{\alpha/2}$ = 1.96
For 80% power	z_{β} = 0.84
For 90% power	z_{β} = 1.28

35

35

What is the variability for AUC?

- Obtain an estimate from pilot study
- $Var_{subject}$ = AUC × (1-AUC) *(conservative)*

36

36

Sample Size Formula for One-Reader Study

$$N_{dis} = \frac{(z_{\alpha/2} + z_{\beta})^2 \times [\ 2 \times \{Var_{subject} \times (1-r_{subject})\ \}\ ]}{(\theta_{new} - \theta_{old} + \delta)^2}$$

$$N_{dis} = \frac{(1.96 + 0.84)^2 \times [\ 2 \times \{0.8 \times 0.2 \times (1-0.5)\ \}\ ]}{(0.05)^2}$$

≈ 500

37

37

- One-reader study: 1000 subjects!
- How generalizable are the results from a one-reader study?
- Good motivation to consider multi-reader studies
- Tomorrow Jeffrey will discuss Multi-reader, multi-modality studies

38

38

Estimate and construct CI

State statistical hypothesis:
- superiority
- non-inferiority
- equivalence

Plan Statistical Analysis:
- prospective or retrospective
- paired or unpaired
- clustered or 1 obs/subject

Assumptions:
- performance estimates
- variability
- correlations
- set type I and II error rates

Calculate sample size

39

39

Clustered data
(common in imaging studies)

Clustered data = multiple observations/subject

Examples:
- left and right breast / subject
- six colon segments / subject
- unknown # lesions / subject

40

40

Clustered data

- In data analysis and in sample size estimation, do NOT assume independence!
- A conservative approach is to assume only one observation / subject.

41

41

Sample size estimation with clustered data

- Alternatively, take clustered data into account:

$$\text{design effect} = 1 + (s-1) \times r_{cluster}$$

s = # observations / subject

$r_{cluster}$ = correlation between observations in cluster

42

42

Example with clustered data

- Suppose readers will score each breast in a mammography study.

$$\text{Design effect} = 1 + (s-1) \times r_{cluster}$$
$$= 1 + (2-1) \times 0.5 = 1.5$$

- If we needed 100 subjects (if only one observation per subject), then:

100 x 1.5 = 150 breasts needed

so, 75 subjects needed (150 breasts)

43

43