INTRODUCTION TO SAMPLE SIZE CALCULATIONS

Ms Chido Dziva Chikwari; BSc,MSc, PhDc
Biomedical Research and Training Institute
March 12, 2019
Objectives

- 1) Understand the concepts of sample size, power, effect size, and types of error
- 2) Understand the tradeoffs between these different parameters
Sample Size & Power
Why estimate sample size?

**Purpose**

- Justify the cost and scope of the proposed research
- Define your assumptions about the expected effect size and the minimum relevant effect size
- Anticipate challenges (e.g., loss to follow-up)
- *Make sure your research is not doomed to fail (be underpowered to draw conclusions)*
Example: Unbalanced Coin

- Flip a coin 4 times
  - 3 heads, 1 tail
  - Do you believe it is an unbalanced coin?
- Flip a coin 40 times
  - 30 heads, 10 tails
  - Now do you believe it is an unbalanced coin?
- Flip a coin 400 times
  - 300 heads, 100 tails
  - Are you convinced this is an unbalanced coin?
Example: Unbalanced Coin

With a larger sample size, we become more sure that the difference we observe is a true difference and not due to chance
Example: Unbalanced Coin
Example: Unbalanced Coin

The coin is either truly balanced or unbalanced
   This is the TRUTH

When we flip the coin a certain number of times we’ll either conclude that the coin is balanced or unbalanced
   This is what we OBSERVE

Most of the times we’ll conclude correctly (observed = truth), but sometimes we’ll conclude incorrectly (observed ≠ truth) simply due to chance
How willing are we to risk making the **wrong** conclusion?

Increasing sample size has **tradeoffs**…

- **Greater** power and certainty
- **Higher** cost
- **More** participants at risk

How do we **quantify** this uncertainty?
P-values

- p-value tells us what the probability is that we would see the observed difference due to chance EVEN IF there was truly no difference
  - How likely is it that the coin toss would show 30 heads and 10 tails if it was truly a balanced coin?

- p-value decreases with greater number of subjects

Does a p-value tell us how strong an association is?
Measures of association

- Calculated to explain whether, in which direction and how strongly an exposure is associated with an outcome
  - RR=1 means the exposure is not associated with the outcome
  - RR<1 means the exposure is negatively associated with the outcome and decreases risk
  - RR>1 means the exposure is positively associated with the outcome and increases risk

- RR = 0.8
  - “Subjects in the exposed group were 80% as likely / 20% less likely to have the outcome”

- RR = 1.5
  - “Subjects in the exposed group were 1.5 times as likely / 50% more likely to have the outcome.”
My friend wants to improve her chances of getting pregnant; she can take one of two medications:

- **Medication 1**: Increases her chance of getting pregnant by a factor of 1.1 and has a p-value of <0.0001
- **Medication 2**: Increases her chance of getting pregnant by a factor of 4.5 and has a p-value of 0.1

Which medication would you choose?
What is Statistical Power?

- Power is a measure of our ability to convincingly **show** a difference when there **truly is** a difference

  - (REMEMBER: p-value is a measure of how often we would **observe** a difference when there **truly was no** difference)

- If I asked you to find out whether a coin was balanced but I only let you toss the coin 4 times, would you be able to make a convincing conclusion?

- There is no point in doing a study unless we know that if there is a true difference, we will be able to **show** it
Components of Statistical Power

Power is a function of four parameters

- **Effect size**
  - How **big** is the effect of your intervention?

- **Sample size**
  - How **many people** (or observations) do you have?

- **Probability of false negative (β error)**
  - \((1 - \beta)\) is **power**

- **Probability of false positive (α error)**
Components of Statistical Power

Power is a function of four parameters:

- **Effect size**
  - How *unbalanced* is your coin?
  - What if it only showed heads 60% of the time? (6/10)

- **Sample size**
  - How *many times* do you need to flip your coin?

- **Probability of false negative (β error)**
  - How willing are you to conclude that the coin is *BALANCED* when it is truly *UNBALANCED*?

- **Probability of false positive (α error)**
  - How willing are you to conclude that the coin is *UNBALANCED* when it is truly *BALANCED*?
Alpha and Beta?

Prisoner Example

<table>
<thead>
<tr>
<th>VERDICT</th>
<th>TRUTH</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Guilty</td>
<td>Innocent</td>
<td></td>
</tr>
<tr>
<td>Prison</td>
<td>OK</td>
<td>Problem</td>
<td></td>
</tr>
<tr>
<td>No prison</td>
<td>Problem</td>
<td>OK</td>
<td></td>
</tr>
</tbody>
</table>

Which is worse?
Problem A in RED or Problem B in GREEN?
Alpha and Beta?

Study Example

<table>
<thead>
<tr>
<th>TRUTH</th>
<th>Effect</th>
<th>No Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effect</td>
<td>OK</td>
<td>Problem (α error)</td>
</tr>
<tr>
<td>No Effect</td>
<td>Problem (β error)</td>
<td>OK</td>
</tr>
</tbody>
</table>

OBSERVED

Which is worse?
Problem A “alpha – Type I error” and Problem B “beta- Type II error”
Alpha usually 0.05 and beta usually 0.2 or 0.1

Willing to accept that 20% of the time we won’t be able to convincingly show that there is a difference, even when there truly is a difference; and 5% of the time we will conclude that there is a difference, even when there truly is NO difference.
What to solve for?

Power is a function of four parameters

- Effect size
- Sample size
- Power (1-\( \beta \))
- Probability of false positive (\( \alpha \) error)

Like in algebra, set 3 equal to real numbers and solve for the 4th
What to solve for?

I have…

- Effect size
- Power \((1-\beta)\)
- Probability of false positive \((\alpha\) error)  

What do I solve for?

- Sample size
What to solve for?

I have...

- Sample size
- Effect size
- Probability of false positive (α error)

What do I solve for?

- Power (1-β)
What to solve for?

I have…

- Sample size
- Power (1-\(\beta\))
- Probability of false positive (\(\alpha\) error)

What do I solve for?

- Effect size
How are these concepts connected?

There are tradeoffs…

- ↓ detectable effect size
- ↓ chance
- ↑ power

- ↑ sample size
- ↑ cost
Working with a statistician

- Statistician needs 3 of the 4 values:
  - Effect size
  - Sample size
  - Power (1-\(\beta\))
  - Probability of false positive (\(\alpha\) error)
Working with a statistician

- Statistician needs 3 of the 4 values:
  - Effect size – estimate from the literature or decide what effect size would be clinically meaningful
  - Sample size – may have a fixed number of participants or a limited budget
  - Power \((1-\beta)\) – usually 80% (sometimes 90% in clinical studies)
  - Probability of false positive \((\alpha \text{ error})\) – usually 5%
Working with a statistician

Example: I want to test whether initiating HIV-positive children on a more palatable ART regimen reduces the chances of virologic failure after 6 months of treatment

- Sample size: UNKNOWN
- Effect size: Unsure, but anything less than a 10% reduction in virologic failure would be clinically irrelevant. Current regimens have probability of failure of 30%
- Power: 80%
- Alpha: 5%

Statistician will tell you how many people you would need to enroll to detect a difference as small as a 10% reduction in virologic failure from the baseline of 30% failure
Example: I want to test whether viral loads of HIV-positive children who received an intensive ART adherence intervention are higher or lower than children who received standard of care adherence counseling. However, I only have funds to process a total of 50 samples

- Sample size: 50 samples, 25 with the intervention, 25 SoC
- Effect size: Unknown
- Power: 80%
- Alpha: 5%

Statistician will tell you how much of a difference in viral load you would be able to detect with 50 samples.
Calculating your own sample size

- Manual formulas or statistical software both valid options
- Most important to understand the inputs well
- Many statistical software options (non-exhaustive list)
  - Stata, SPSS, R: all flexible and powerful, require knowledge of coding
  - EpiInfo (StatCalc): free and user-friendly, also have web-based and mobile-based options
  - Power Sample Size (Vanderbilt University) is free and easy to use, does not require coding:
    http://biostat.mc.vanderbilt.edu/wiki/Main/PowerSampleSize
. power twoproportions 0.8 0.9, alpha (0.05) power (0.80)

Performing iteration ... 

Estimated sample sizes for a two-sample proportions test
Pearson's chi-squared test
Ho: p2 = p1 versus Ha: p2 != p1

Study parameters:

alpha = 0.0500
power = 0.8000
delta = 0.1000 (difference)
p1 = 0.8000
p2 = 0.9000

Estimated sample sizes:

N = 398
N per group = 199
. power twoproportions 0.6 0.9, alpha (0.05) power (0.80)

Performing iteration ...

Estimated sample sizes for a two-sample proportions test
Pearson's chi-squared test
Ho: p2 = p1   versus   Ha: p2  !=  p1

Study parameters:

  alpha =  0.0500
  power =  0.8000
  delta =  0.3000  (difference)
  p1 =  0.6000
  p2 =  0.9000

Estimated sample sizes:

  N =  64
  N per group =  32
. `power twomeans 2.5 2.75, alpha (0.05) power (0.80)`

Performing iteration ...

Estimated sample sizes for a two-sample means test
`t test` assuming sd1 = sd2 = sd
Ho: m2 = m1 versus Ha: m2 != m1

Study parameters:

alpha = 0.0500
power = 0.8000
delta = 0.2500
  m1 = 2.5000
  m2 = 2.7500
  sd = 1.0000

Estimated sample sizes:

N = 506
N per group = 253
Using **YOUR** study question, imagine you are either working with a biostatistician to do power calculations OR calculating your own sample size. Define 3 of the following 4 parameters and write a 1 sentence description of what the biostatistician or software will tell you.

- Effect size
- Power
- Sample size
- Alpha error
Summary
Components of statistical power:
- Effect size
- Sample size
- Power
- Apha error
Special thanks to:

Dr. Anjuli Wagner, MPH, PhD
Department of Global Health, University of Washington