BIOSTATISTICS WORKSHOP: SAMPLE SIZE & POWER

Sub-Saharan Africa CFAR meeting
July 18, 2016
Durban, South Africa

Memory loss and Dementia in HIV

- Does HIV infection accelerate onset of memory loss at advanced ages?
- What do we need to determine in planning this study?
  - Study design?
  - Study participants?
  - Endpoints? Measures of memory loss?
  - Covariates / possible confounders to collect?
  - Statistical Analysis Plan (SAP)
- Who should be involved in the process of planning a study?

Memory loss and Dementia in HIV

- Does HIV infection accelerate onset of memory loss at advanced ages?

- We want to recruit a sample of HIV+ & HIV- individuals between ages of 55 and 65 and test their memory
  - Primary endpoint: Memory as a continuous measure where lower values indicate worse memory
  - Secondary endpoint: Self-assessed memory impairment (“Do you feel your memory today is worse than three years ago?”)
- Confounders
  - Age, medication use & duration, age at HIV onset?
- Statistical Analysis Plan?
- How many participants do you need to see a meaningful difference (if one exists)?

Sample Size Calculations

- Before beginning a study you want to determine how many subjects you will need to enroll
  - To see the desired / expected effect
  - To have a high probability that that effect is statistically significant (assuming the effect exists)

- Power calculations can be used to:
  - Determine the sample size needed
  - Determine the power given a fixed or maximum sample size
  - Determine the detectable effect size given a sample size and power

So, do we really need to do this?

- **Yes!!!!**
- If a sample size isn't large enough,
  - we may conclude a null result (even if there truly is an effect) due to a lack of statistical power (type II error)
- If sample size is too large,
  - we have wasted valuable resources (time, $, etc.)

Our decisions & mistakes

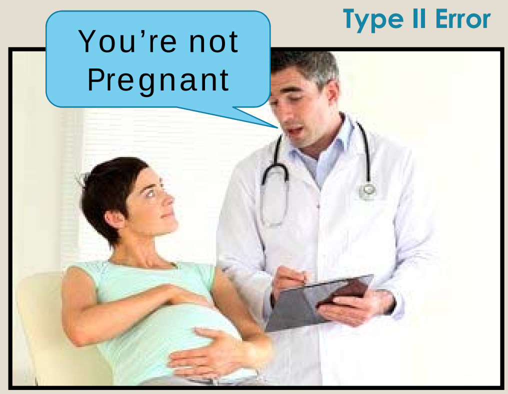
		Reality!	
		H_0 is true There is not a difference	H_0 is False There is a difference
Conclusion	Do Not Reject H_0 There is not a difference	Correct ☺	Type II Error
	Reject H_0 There is a difference	Type I Error	Correct ☺

P(type I error) = P(rejecting H_0 | H_0 is true)= **α**

P(type II error) = P(not rejecting H_0 | H_0 is false)= **β**

P(rejecting H_0 | H_0 is false) = **1-β = POWER**

Type I vs. Type II error



Type I vs Type II error

- Type I error: Significance Level **Societal Risk**
  - α = P(rejecting H_0 | H_0 is true)
  - Incorrectly concluding that there is a difference when there truly is not a difference (concluding a drug works, when it in fact does not)
  - False Positive
  - Typically set at 5% overall

- Type II Error: Power **Institutional Risk**
  - P(not rejecting H_0 | H_0 if false)
  - Correctly concluding that an effect exists when it does (finding a drug works when in fact it does)
  - True Positive
  - Expressed as a %, typical values 80% & 90%

Memory loss and Dementia in HIV

- Does HIV infection accelerate onset of memory loss at advanced ages?
- We want to recruit a sample of HIV+ & HIV- individuals between ages of 55 and 65 and test their memory
  - **Primary endpoint: Memory as a continuous measure where lower values indicate worse memory**
  - Secondary endpoint: Self-assessed memory impairment ("Do you feel your memory today is worse than three years ago?")

Sample Size Calculation

- What you need
  - Estimate of expected effect size
  - Estimate of expected variability
  - Significance level
    - Typically $\alpha = 0.05$
    - Take into account # of endpoints and tests
  - Power – Probability of finding a significant effect given that effect exists

SS Calculation: Effect Size

- Depends on analysis to be performed
  - Difference in means? OR? RR?
- Clinically Meaningful / Relevant Difference
- Realistic, but reasonable

- How to get an estimate
  - Previous literature
  - Pilot Study
  - Clinically Meaningful
  - Increments of Standard Error

SS Calculation: Variability

- Usually measured as Standard Deviation or proportion expected in each group
- Clinically Meaningful / Relevant Difference
- Realistic, but reasonable

- How to get an estimate
  - Previous literature
  - Pilot Study
  - Tricks?
    - (max – min) / 4
  - Err on side of over-estimate

Memory loss and Dementia in HIV

- Does HIV infection accelerate onset of memory loss at advanced ages?

- Compare HIV+ and HIV- individuals on a continuous, normally-distributed memory score
- 2-sample t-test
- We find previous literature using this measure with HIV- individuals and they reported, for 55-65 years olds, a mean of 20 with a standard deviation of 5
- We want to see a difference of 1 SD (5 units) between the groups

- How many subjects do we need to recruit in each group to see a difference of 5 units?

R Programming

- Free Software program available to download
- www.r-project.org

- I will show you some very simple code for straight-forward sample size calculations

- Many more examples can be found just a google away!

Sample Size in R

```
>
> power.t.test(delta=5, sd=5, power=0.8, sig.level=0.05)
```

R function

Parameters we need to add

Delta = difference in means

Sample Size in R

```
>
> power.t.test(delta=5, sd=5, power=0.8, sig.level=0.05)

     Two-sample t test power calculation 

              n = 16.71477
          delta = 5
             sd = 5
      sig.level = 0.05
          power = 0.8
    alternative = two.sided

NOTE: n is number in *each* group
```

Need 17 HIV+ and 17 HIV- to find a difference in 5 units in memory score

Sample Size

- So what happens if we enroll 17 + 17 people and the true difference is actually less than 5 units?
- We consult the neuropsych people and they say that a difference in just 2 units would be considered clinically meaningful
- How many subjects do we need to recruit in each group to see a difference of 2 units?

More or less than 34?

Sample Size in R

```
>
> power.t.test(delta=5, sd=5, power=0.8, sig.level=0.05)
```

What do we have to change?

```
> power.t.test(delta=2, sd=5, power=0.8, sig.level=0.05)
```

Sample Size in R

```
> power.t.test(delta=2, sd=5, power=0.8, sig.level=0.05)

     Two-sample t test power calculation

              n = 99.08057
          delta = 2
             sd = 5
      sig.level = 0.05
          power = 0.8
    alternative = two.sided

NOTE: n is number in *each* group
```

How many do we need in each group to see a difference in 2 units in memory score?

Sample Size

- We are gambling people, so we want to up our probability of finding significance (if effect exists) so we will increase power to 90%
- How many subjects do we need to recruit in each group to see a difference of 2 units at 90% power?

More or less than 200?

Sample Size in R

```
> power.t.test(delta=2, sd=5, power=0.8, sig.level=0.05)
```

What do we have to change?

```
> power.t.test(delta=2, sd=5, power=0.9, sig.level=0.05)
```

Sample Size in R

```
> power.t.test(delta=2, sd=5, power=0.9, sig.level=0.05)

     Two-sample t test power calculation

              n = 132.3106
          delta = 2
             sd = 5
      sig.level = 0.05
          power = 0.9
    alternative = two.sided

NOTE: n is number in *each* group
```

How many do we need in each group to see a difference in 2 units in memory score at 90% power?

Sample Size

- Smaller differences ↑ n
- Larger standard deviations ↑ n
- More power ↑ n
- Stronger type I error control (smaller α) ↑ n
- More narrow CI ↑ n

Calculating power from n

- Sometimes you have a fixed n and want to calculate power to find a particular effect size
- Sometimes you reach the end of your study, fail to reject H_0 and want to see if you had enough power to find significance for the effect size you have
  - “Post-hoc Power Calculation”

Power in R

```
> power.t.test(delta=2, sd=5, power=0.9, sig.level=0.05)
```

What do we have to change?

```
> power.t.test(delta=2, sd=5, n=99, sig.level=0.05)
```

Whatever you don't indicate is what R calculates

Power in R

```
> power.t.test(delta=2, sd=5, n=99, sig.level=0.05)

     Two-sample t test power calculation

              n = 99
          delta = 2
             sd = 5
      sig.level = 0.05
          power = 0.7996777
    alternative = two.sided

NOTE: n is number in *each* group
```

Do we reach 80% power with 99 in each group?

Power in R

```
> power.t.test(delta=2, sd=5, n=100, sig.level=0.05)

     Two-sample t test power calculation

              n = 100
          delta = 2
             sd = 5
      sig.level = 0.05
          power = 0.8036466
    alternative = two.sided

NOTE: n is number in *each* group
```

What does 80% power really mean?

Is there something magical about 80% power?

Quick Re-Cap

- Does HIV infection accelerate onset of memory loss at advanced ages?
- Compare HIV+ and HIV- individuals on a continuous, normally-distributed memory score
- 2-sample t-test
- We need to enroll 100 HIV+ and 100 HIV- individuals into study to see a difference in means of 2 units at 80% power
- At 90% power we need to enroll 133 HIV+ and HIV- individuals into the study

Hypothetical Study Question

- Does HIV infection accelerate onset of memory loss at advanced ages?

- We want to recruit a sample of HIV+ & HIV- individuals between ages of 55 and 65 and test their memory
  - Primary endpoint: Memory as a continuous measure where lower values indicate worse memory
  - **Secondary endpoint: Self-assessed memory impairment ("Do you feel your memory today is worse than three years ago?")**

Memory loss and Dementia in HIV

- Does HIV infection accelerate onset of memory loss at advanced ages?

- Compare HIV+ and HIV- individuals on a binary variable
- Chi-Square test, our effect measure is OR (case-control study)
- We find previous literature using this measure with HIV- individuals and 15% reported having experienced worse memory than 3 years earlier.

- We think that HIV+ people will have 2xs the odds of reporting worse memory

- How many subjects do we need to recruit in each group to see an odds ratio = 2?

Risk (p) vs. Odds (o)

$$o = \frac{p}{1-p} \qquad p = \frac{o}{1+o}$$

$$odds\ ratio\ (OR) = \frac{o_1}{o_2} = \frac{p_1 / (1-p_1)}{p_2 / (1-p_2)}$$

We find previous literature using this measure with HIV- individuals and 15% reported having experienced worse memory than 3 years earlier.

In this case 15% is a proportion or 'risk' and we need to calculate an OR

For simple R sample size calculations we need p_1 & p_2

We have p_2 (15%) & OR, need to estimate p_1

$$p_1 = \frac{OR * p_2}{1 - p_2 + OR * p_2}$$

Risk (p) vs. Odds (o)

$$o = \frac{p}{1-p} \qquad p = \frac{o}{1+o}$$

$$odds\ ratio\ (OR) = \frac{o_{HIV+}}{o_{HIV-}} = \frac{p_{HIV+} / (1-p_{HIV+})}{p_{HIV-} / (1-p_{HIV-})}$$

We find previous literature using this measure with HIV- individuals and 15% reported having experienced worse memory than 3 years earlier.

In this case 15% is a proportion or 'risk' and we need to calculate an OR

For simple R sample size calculations we need p_1 & p_2

We have p_2 (15%) & OR, need to estimate p_1

$$p_{HIV+} = \frac{OR * p_{HIV-}}{1 - p_{HIV-} + OR * p_{HIV-}}$$

Risk (p) vs. Odds (o)

$$o = \frac{p}{1-p} \qquad p = \frac{o}{1+o}$$

$$OR = 2.0 = \frac{p_{HIV+} / (1-p_{HIV+})}{0.15 / (1-0.15)}$$

We find previous literature using this measure with HIV- individuals and 15% reported having experienced worse memory than 3 years earlier.

In this case 15% is a proportion or 'risk' and we need to calculate an OR

For simple R sample size calculations we need p_1 & p_2

We have p_2 (15%) & OR, need to estimate p_1

$$p_{HIV+} = \frac{2.0 * 0.15}{1 - 0.15 + 2.0 * 0.15} = 0.26$$

Sample Size in R

```
> power.prop.test(p1=0.26, p2=0.15, power=0.8)
```

R function: `power.prop.test` — Parameters we need to add: `p1=0.26, p2=0.15, power=0.8`

Sample Size in R

```
> power.prop.test(p1=0.26, p2=0.15, power=0.8)

     Two-sample comparison of proportions power calculation

              n = 210.2505
             p1 = 0.26
             p2 = 0.15
      sig.level = 0.05
          power = 0.8
    alternative = two.sided

NOTE: n is number in *each* group
```

So see an OR=2, at 80% power and a proportion in the HIV- group = 15% we will need 211 in each group

Memory loss and Dementia in HIV

- Does HIV infection accelerate onset of memory loss at advanced ages?
- Compare HIV+ and HIV- individuals on a binary variable
- Chi-Square test, our effect measure is OR (case-control study)
- We decide we want to study 65-75 year olds. In that population 30% of HIV- individuals report experiencing worse memory than 3 years earlier.
- We still think that HIV+ people will have 2xs the odds of reporting worse memory
- How many subjects do we need to recruit in each group to see an odds ratio = 2?

Sample Size in R

```
> power.prop.test(p1=0.26, p2=0.15, power=0.8)
```

What do we have to change?

Risk (p) vs. Odds (o)

$$o = \frac{p}{1-p} \qquad p = \frac{o}{1+o} \qquad OR = 2.0 = \frac{p_{HIV+} / (1-p_{HIV+})}{0.35 / (1-0.35)}$$

We find previous literature using this measure with HIV- individuals and 15% reported having experienced worse memory than 3 years earlier.

In this case 15% is a proportion or 'risk' and we need to calculate an OR

For simple R sample size calculations we need p_1 & p_2

We have p_2 (15%) & OR, need to estimate p_1

$$p_{HIV+} = \frac{2.0 * 0.35}{1 - 0.35 + 2.0 * 0.35} = 0.52$$

Sample Size in R

```
> power.prop.test(p1=0.26, p2=0.15, power=0.8)
```

What do we have to change?

```
> power.prop.test(p1=0.52, p2=0.35, power=0.8)
```

Sample Size in R

```
> power.prop.test(p1=0.52, p2=0.35, power=0.8)

     Two-sample comparison of proportions power calculation

              n = 132.3138
             p1 = 0.52
             p2 = 0.35
      sig.level = 0.05
          power = 0.8
    alternative = two.sided

NOTE: n is number in *each* group
```

So see an OR=2, at 80% power and a proportion in the HIV-group = 35% we will need 133in each group

Sample Size presentation

◦ Not uncommon to present multiple possibilities in a power/sample size section of a grant.
◦ Vary effect size, power and variability
◦ Do NOT vary significance level!

HIV+ (p)	HIV- (p)	OR	Power	n per group
0.26	0.15	2.0	80%	211
0.26	0.15	2.0	90%	281
0.35	0.15	3.0	80%	73
0.35	0.15	3.0	90%	97

Sample Size: Notes

◦ Calculation (once you have the inputs) is relatively simple, but estimation of ES can be difficult
◦ Important to be conservative but maintain reason when estimating parameters
◦ Small changes in some parameters may have a large effect on the power
◦ In the end, it's often a balancing act

◦ Take into account the # of tests and endpoints you have.
◦ Adjust alpha (sig.level in R) to control for multiple comparisons

Memory loss and Dementia in HIV

- Does HIV infection accelerate onset of memory loss at advanced ages?
- What if we wanted to follow participants and measure change in memory over time.
- Longitudinal study
  - Visit them at baseline, year 1, year 2 and year 3
- At the end of 3 years we ask them "Do you feel your have worse memory than 3 years ago?"
- Does the change in study design effect our sample size calculation?

Longitudinal Study

- How many people do we need to enroll in each clinic (at 80% power) to see an OR=2.0?

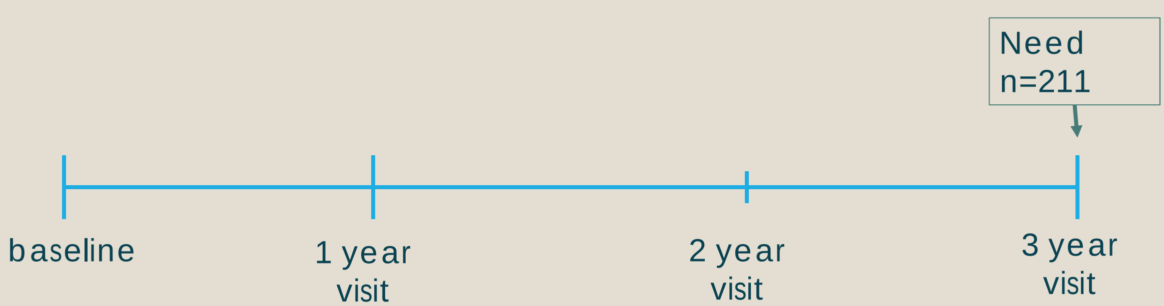
```
> power.prop.test(p1=0.26, p2=0.15, power=0.8)

     Two-sample comparison of proportions power calculation

              n = 210.2505
             p1 = 0.26
             p2 = 0.15
      sig.level = 0.05
          power = 0.8
    alternative = two.sided

NOTE: n is number in *each* group
```

Longitudinal Study

- For prospective studies
- Need to take into account 'drop-outs'
  - Say you enroll 211 people at baseline
  - Can you realistically expect to see 211 people at 1 year follow-up?
  - What about at 3 years?
- Sample Size calculation is for number needed at END of study
- So you need an additional estimate for expected "loss to follow-up"

Longitudinal Study

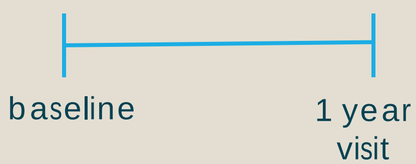
- How many people do we need to enroll in each clinic (at 80% power) to see an OR=2.0?



Longitudinal Study

- Start more simple
  - Let's say it was a 1 year study
  - We expect to lose 10%

Need n=211



If we expect to lose 10%, that means that at 1 month we expect to have 90% of what we started with

$$\frac{90}{100} = \frac{211}{X}$$

$$X = \frac{100*211}{90}$$

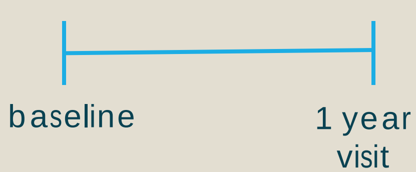
$$X = 234.4$$

$$235 \textit{ subjects per group}$$

longitudinal: another option

- Start more simple
  - Let's say it was a 1 month trail
  - We expect to lose 10%

Need n=211



If we expect to lose 10%, that means that at 1 month we expect to have 90% of what we started with

$$0.90 = \frac{211}{X}$$

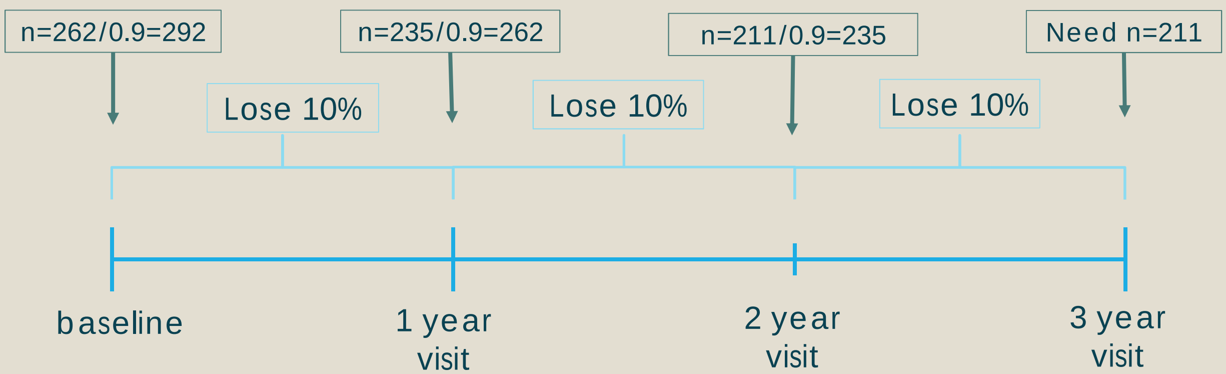
$$X = \frac{211}{0.9}$$

$$X = 234.4$$

$$235 \textit{ subjects per group}$$

Longitudinal Study

- Now do that 3 times



Longitudinal: in 1 step?

Lose some people due to rounding

Sample Size: Common PitFalls

- Drop outs
- Secondary Endpoints
- Multiplicity
- Recognizing Futility
- Choosing the wrong endpoint
- Massaging the parameters to get 80% power will not help you in the end!!!

Sample Size: Notes

- Calculation (once you have the inputs) is relatively simple, but estimation of ES can be difficult
- Important to be conservative but maintain reason when estimating parameters
- Small changes in some parameters may have a large effect on the power
- In the end, it's often a balancing act

Sample Size: Summary

- Perform sample size calculations during the design phase of your research
- Ensure that you will have enough power to detect a difference if one exists
- Absence of evidence of an effect is not the same as evidence of absence of an effect (power may be too low)
- Know when to consult a statistician!

To consult the statistician after an experiment is finished is often merely to ask him to conduct a post-mortem examination. He can perhaps say what the experiment died of.

R.A Fisher (1890-1962)

Sample Size: Software

- GraphPad Prism
  - researcher user friendly
  - point and click
- Free online tools (genetics based)
  - Shaun Purcell: http://pngu.mgh.harvard.edu/~purcell/gpc/
  - Quanto: http://hydra.usc.edu/gxe/
  - Harvard/MGH: http://hedwig.mgh.harvard.edu/sample_size/size.html
  - Others out there… but beware!
  - R & R Studio www.r-project.org

QUESTIONS?