

# Session 2

## Sample size calculations

1st April 2019

Chan Yiong Huak  
Head, Biostatistics Unit  
Yong Loo Lin School of Medicine  
National University of Singapore









Are you a p-value worshipper?

What does statistical significance mean ?

p-value is affected by sample size

$n \uparrow$     p-value  $\downarrow$

# Statistical vs Clinical significance

Clinical Significance	Statistical Significance	Meaning?
		Results reproducible
		Sample size small
		Sample size large
		Forget about it!

## How to report p-values

Never report  $p = 0.000$

Small p-values :  $p < 0.001$  (or  $p < 0.0005$ )

$p < 0.05$

$p < 0.01$

Not  $p < 0.2$

$p = 0.2$

# Sample size

Comparing 2 groups

## Sample size

Is a sample size of 20 too small?

## Sample size

For 80% power & a 2-sided test 5%, if 80% of the subjects on treatment A had a success compared to 20% on treatment B, then only 10 subjects on each arm will achieve a statistically significant result!

Total sample size = 20!



## Sample size

Is a sample size of 1000 large?

## Sample size

For 80% power & a 2-sided test 5%, to increase the success rate of 1% on treatment A to 2% on treatment B, then will need 2400 subjects on each arm to achieve a statistically significant result!

Thus 1000 subjects not enough!

## Sample size

Epidemiological studies

No power calculations  
(except for case-control)

Otherwise it is the 95% CI on the  
prevalence/incidence of interest  
(at least  $n = 400$  or more)

Sample size is **NOT A MAGIC NUMBER!!!!**

It is dependent on the PRIMARY outcome that you want to show a CLINICAL difference.

A number to calculate (part of) the amount required for grant application

A number to estimate your accrual rate or data collection

## Some definitions

The null hypothesis is set up to directly counter the research question. The philosophical argument is that it is harder to prove something to be true, but easier to prove it to be false

Intention to Treat (*ITT*)

*Vs*

Per Protocol (*PP*)

# HYPOTHESIS TESTING

## 1-Sided vs 2-Sided Tests

The two-sided test is more conservative and should be used when in doubt

## Type I Error

Rejecting the Null Hypothesis when it is true, sometimes known as the false-positive error (the probability of incorrectly declaring a treatment difference when no difference exists)



## P- Value

The probability of obtaining a result as extreme as or more extreme than the one observed, *if* the null hypothesis is true; or sometimes defined as the probability that the observed result is due to chance alone

## Type II Error

Accept the Null Hypothesis when it is false  
– false negative error (the probability of incorrectly failing to declare a treatment difference when a difference exists)

# Power

1 - Type II Error

The probability of rejecting  $H_0$  when it is actually false

# Sample Size – Power Calculations

Primary response - Must be objective

Statistical vs Clinical significance

## Type of Comparison

Superiority	: A better B
Equivalence	: A as good as B
Non-inferior	: A just slightly ‘worst’ than B

# SAMPLE-SIZE CALCULATIONS

## Superiority Studies

### Case 1

You postulate that 65% of the subjects on Trt A have success (yes/no) outcome compared to 30% on Trt B.

What sample size is required?

# SAMPLE-SIZE CALCULATIONS

5% Significance, 2-sided test  
At 90% Power

65% vs 30%	: 41 per group
vs 35%	: 57 per group
vs 50%	: 227 per group
vs 55%	: 503 per group

# SAMPLE-SIZE CALCULATIONS

Sample Size increases with

- Bigger Power
- Smaller difference

# SAMPLE-SIZE CALCULATIONS

## Case 2

(assume nearly all pop-pop)

The mean (sd) of a numerical outcome for Trt A is 500 (400) units compared to 400 (200) units for Trt B.

What sample size is required?



# SAMPLE-SIZE CALCULATIONS

5% Significance, 2-sided test  
Difference = 100, 90% Power

SD = 200	:	86	per group
250	:	100	per group
300	:	191	per group
400	:	338	per group

# SAMPLE-SIZE CALCULATIONS

Sample Size increases with

- Bigger SD

A more realistic sd is obtained using  
 $(\text{max}-\text{min})/4$

## Equivalence Trials

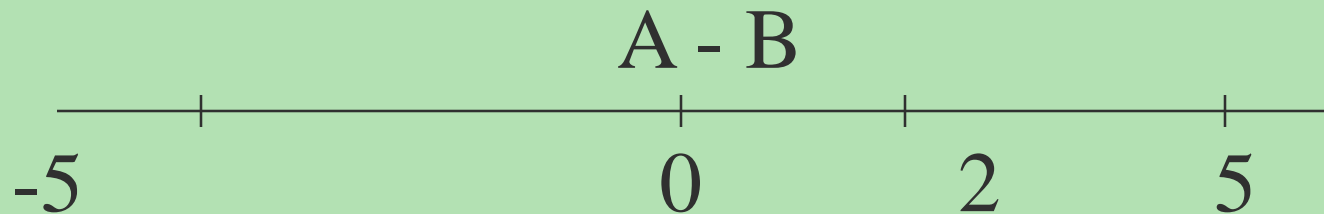
A common ‘conclusion’ from studies which found no differences between 2 treatments is to say that both treatments are ‘equivalent’.

This statement is not valid !

# Equivalence Trials

P-values not used. 95% CI of difference presented

Eg. Declare interval of equivalence is 5%. Say difference between A & B is 2%, then



Equivalence declared 

Equivalence not  declared

95% CI of A - B

# Epidemiology

- Determining the Incidence/Prevalence of a condition
- Finding risk predictors
  - is there a control group?
  - is it a matched control?
- Retrospective vs Prospective

# Epidemiology

- Sampling plan
- Sample Size
  - No power calculations
  - Possible to determine the sample size to ‘capture’ the incidence/prevalence with a certain confidence interval
  - The ‘bigger’ the sample size the better the precision

The End

Thank You