RESEARCH METHODS
Getting the numbers right: sample size in clinical trials

Andrzej Januszewski MD, PhD, MClinTRes

University of Sydney

THE UNIVERSITY OF SYDNEY

Slides handouts: https://bit.ly/sampsize-handouts

1

Clinical trials - basics

- **Research studies that test how well new drugs or treatments work in people.**
- **Goals:**
  - Assess efficacy.
  - Monitor safety and side effects.
  - Determine optimal dosage.
- **Why pharmacists care:**
  - Involvement in medication management during trials.
  - Understanding evidence behind medicines they dispense.
  - Pharmacovigilance and patient education.

Slides handouts: https://bit.ly/sampsize-handouts

2

Design and purpose of trials

- **Study Designs:**
  - Randomised Controlled Trial (RCT) – most rigorous.
  - Parallel vs. Crossover.
  - Blinded (single/double) vs. Open-label.
- **Purpose:**
  - Superiority: Is new drug better?
  - Non-Inferiority: Is new drug not worse?
  - Equivalence: Is new drug the same?

Slides handouts: https://bit.ly/sampsize-handouts

3

Why is having the right sample size so important?

- **Large enough to provide reliable answer to the scientific question**
- **Too small**
  - → **waste of time to do research**
  - → **possibly unethical**
- **Too large**
  - → **waste of money**
  - → **difficult to implement**

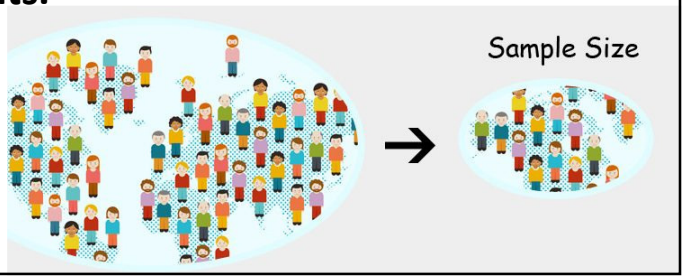
Understanding sample size helps evaluate the quality of trial evidence behind medicines.

Slides handouts: https://bit.ly/sampsize-handouts

4

Sample size

- Only an estimate, based on series of assumptions.
- Often based on very small prior data.
- Published results might be optimistic (publication bias).
- Subjects might be from different populations, have different event definitions / end-points.

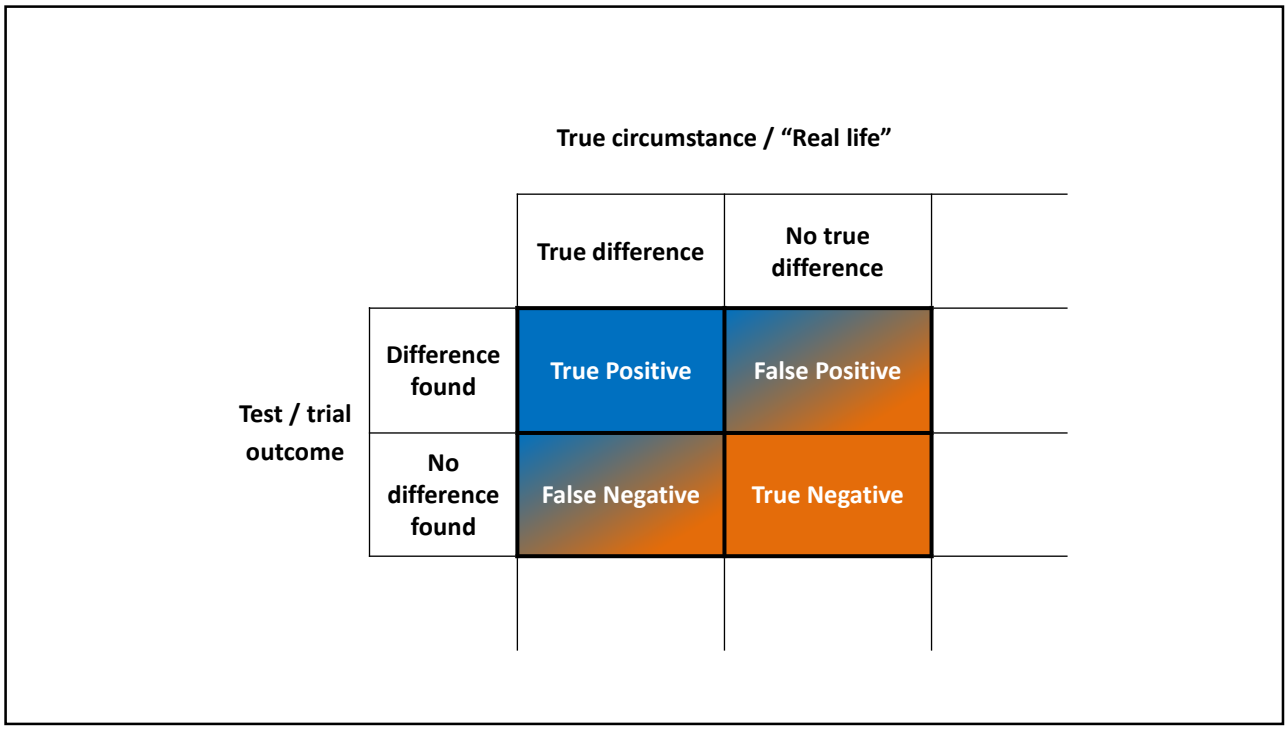


5

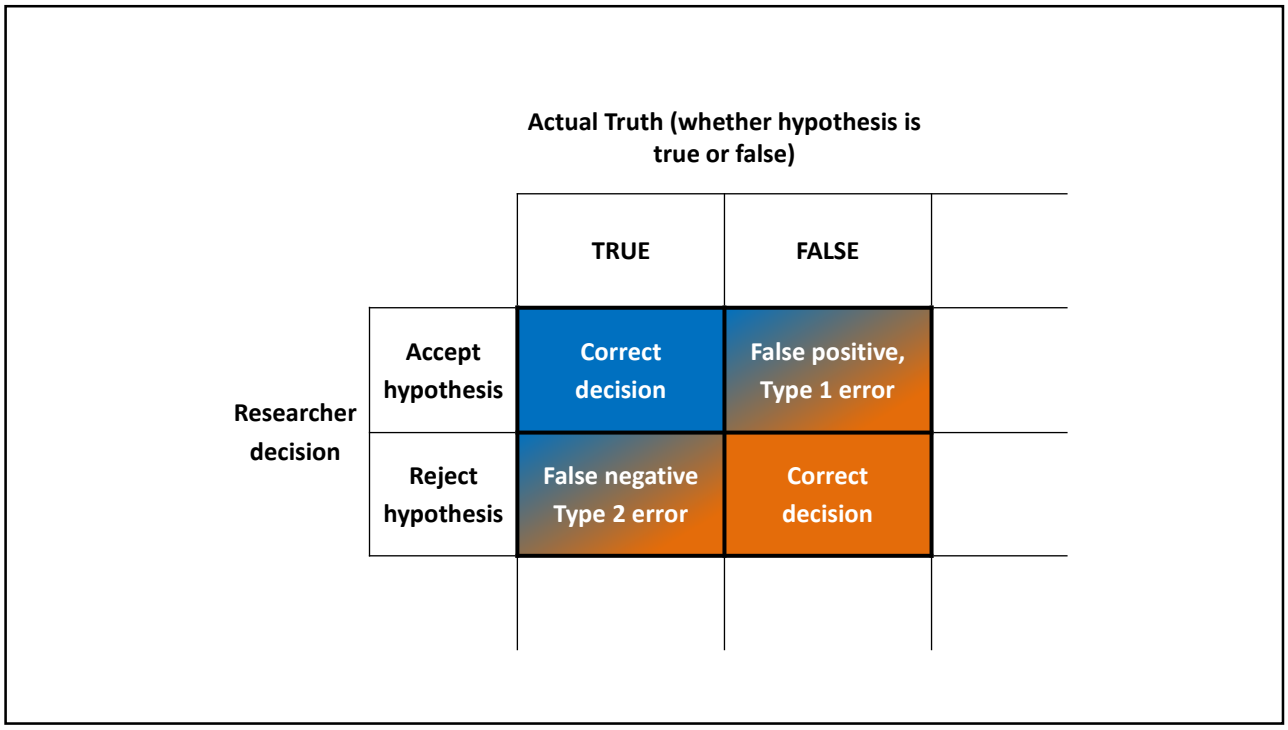
Information to consider for sample size calculations

- What is the purpose of the study?
- What outcome will be measured?
- What outcome rate is expected (% of patients finishing trial)?
- What difference is expected?
- What attrition to the study is expected (no. of patients adhering)?
- What are the cost constraints?

6


True circumstance / "Real life"
True difference
No true difference
Test / trial outcome
Difference found
True Positive
False Positive
No difference found
False Negative
True Negative

7


Actual Truth (whether hypothesis is true or false)
TRUE
FALSE
Researcher decision
Accept hypothesis
Correct decision
False positive, Type 1 error
Reject hypothesis
False negative Type 2 error
Correct decision

8

What is a False Positive result?

(Type I error - α)

- **It is when a difference is declared even though there is no true difference.**
- **The p-value is the probability of a positive result by chance alone. The smaller the set p value, the less chance of declaring a False Positive result.**
- **The smaller the p value, the less is the likelihood that the result seen is due to chance.**

9

What is a False Negative result?

(Type II error – β)

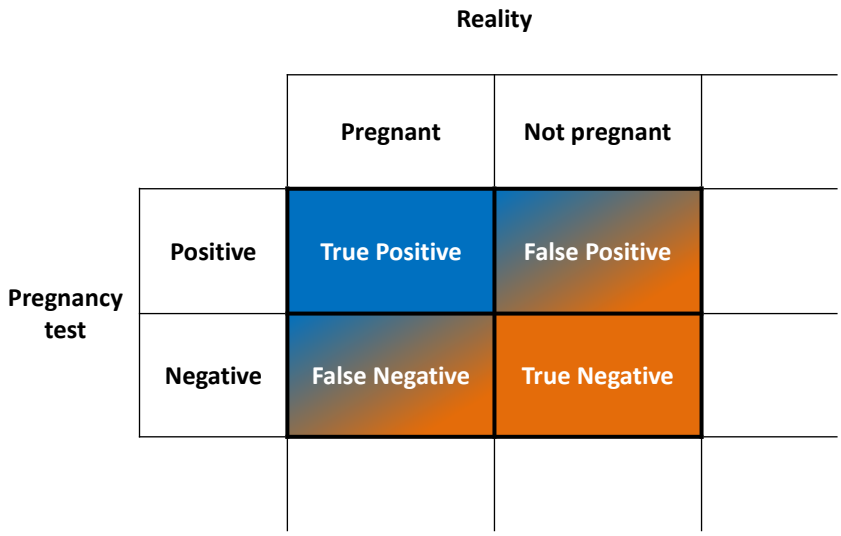
- **It is when no difference is declared even though there is a true difference.**
- **The β is the probability that an experiment will yield a "not statistically significant" result by missing the true effect.**
- **The smaller the β the less risk of a False Negative result.**

10

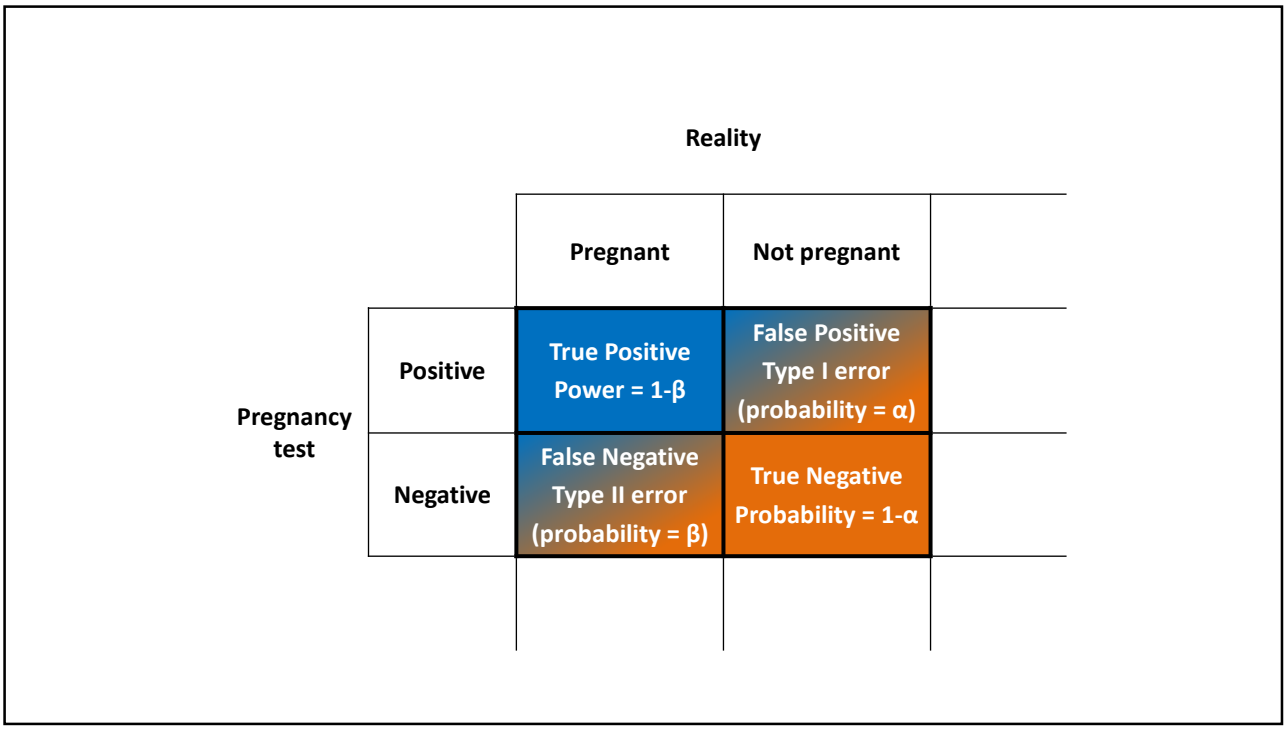
Power = 1-β

- **Statistical power is the likelihood that a study will detect an effect when there is an effect there to be detected.**
- **Probability to achieve statistical significance.**
- **If statistical power is high, the probability of making a Type II error (i.e. concluding there is no effect when, in fact, there is one) goes down.**

11



12



13

A Study Design Challenge

- To calculate sample size that minimises the chance of False Positive and/or False Negative results *within available resources*.

- Minimising False Positive: small p value
- Minimising False Negative: high power

14

The choice of α and β based on:

- **The medical, legal and practical consequences of the type I and type II error(s).**
- **Prior plausibility of the tested hypothesis.**
- **The desired impact of the study results.**

15

The choice of α and β

- **α=0.05 and β=0.20 – most standard for confirmatory trials.**
- **α=0.10 and β=0.20 for preliminary trials that are likely to be replicated.**
- **α=0.01 and β=0.05 for the trials that are unlikely to be replicated.**
- **α = β if both treatments are new, about equal in cost and both are considered safe.**

16

The choice of α and β

- **α > β (sometimes α=0.2 and β=0.1) pilot/exploratory studies if there is no established effective treatment and test treatment is inexpensive, easy to apply and not known to have serious side effects, non-life-threatening context.**
- **α < β (usually α=0.05 and β=0.2) if the control treatment is already widely used, is safe and effective whereas the treatment is new, more expensive and produces serious side effects.**

17

Example of α > β

- **We are testing a new phone app to improve mood in young people.**
  - **It's free, non-invasive, doesn't collect private data, and users can stop anytime.**
  - **There's no current effective intervention.**
- **A Type I error (false positive: thinking the app helps when it doesn't) set low (causing large sample size) might just mean a bit of wasted time or optimism.**
- **But a Type II error (false negative: missing a real benefit) set at e.g. 0.2 (80% power) could mean missed opportunity and not deploying a helpful tool.**
- **It's acceptable – even desirable – to set α > β. <u>We are less worried about false positive and more worried about missing real benefit.</u>**

18

Example 1

- A new treatment to improve glycaemic control,
- Thought to have very few side effects, but quite expensive,
- It is hoped to lower HbA1c by 1.5%,
- Will be marketed if it can lower HbA1c by at least 1.0%,
- In a typical clinic average HbA1c is 8.7% ± 3%.

19

- What is the purpose of the study?
  To improve glycaemic control
- What outcome will be measured?
  HbA1c
- What outcome rate is expected?
  100% subjects (less any losses to follow up)
- What difference is expected?
  1.5% (drop from 8.7% to 7.2%)
- Non-adherence rate?
  0% drop-out expected (100% adherence)
- What are the cost constraints?
  None

20

Formula

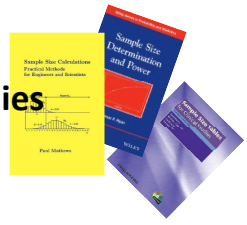
$$\text{Sample Size} = \frac{\left(\frac{1}{\text{p value}} + \text{power}\right)^2 * (\text{standard deviation})^2}{(\text{difference})^2}$$

https://bit.ly/sampsize-formula

21

Resources

- **WEBSITES**
  - http://statpages.org
  - https://bit.ly/sampsize-calc
- **BOOKS**
  - Thomas P. Ryan (ed.) Sample Size Determination and Power (Wiley Series in Probability and Statistics)
  - Paul Mathews (ed.) Sample Size Calculations: Practical Methods for Engineers and Scientists
  - David Machin et al. Sample Size Tables for Clinical Studies



22

A PERFECT Study

- 100% adherence, no losses to follow up
- Sample size for 80% power and 2-sided p≤0.05:

63 per group = 126 for the study

23

https://bit.ly/sampsize-calc

Mean (group 1)	8.7		INPUT
Mean (group 2)	7.2		RESULTS
SD	3		
alpha	0.05	type 1 error rate	
Power (%)	80	1-beta	
1- or 2 tail	2		
group 1 to group 2 ratio	1	ignored if paired design!	
Study design	unpaired		
Drop-outs (%)	0		
Drop-ins (%)	0		
Loss to follow up (%)	0		
n group 1	63		
n group 2	63		
n (per study)	126		
effect size	0.50		

Questions/comments: Andrzej.Januszewski@sydney.edu.au

24

Study power

- 80% power means that there is also a 20% (1 in 5) chance that the study will produce a false negative result.
- 80% is generally regarded as the very minimum acceptable study power.
- In these days, most commercial sponsors would want a much smaller chance of a false negative result i.e. at least 90% power.

25

Varying the study power

(trying to avoid a False Negative result)

- 100% adherence, no losses to follow up
- 2-sided p≤0.05, HbA1c diff. 1.5%, SD=3%

Power	Sample size
80%	
85%	
90%	
95%	

https://bit.ly/sampsize-calc

26

Varying the p value

(trying to avoid a False Positive result)

- 100% adherence, no losses to follow up
- 80% power, HbA1c diff. 1.5%, SD=3%

p-value	Sample size
0.05	
0.01	
0.005	
0.001	

https://bit.ly/sampsize-calc

27

Varying the treatment adherence

(introducing reality)

- 80% power, 2-sided p≤0.05
- HbA1c diff. 1.5%, SD=3%

Adherence	Sample size
100%	
95%	
90%	
80%	

https://bit.ly/sampsize-calc

28

Introducing loss-to-follow up

(introducing reality)

- 80% power, 2-sided p≤0.05
- HbA1c diff. 1.5%, SD=3%

Loss to follow-up	Sample size
0%	
5%	
10%	
20%	

https://bit.ly/sampsize-calc

29

Non-adherence vs. Loss to follow-up

- Non-adherence (or non-compliance) means that the patient is changing treatment she/he is randomised to (drop-out or drop-in).
- If randomised to active treatment and stopping it = switching to placebo (and vice versa).
- Intention To Treat (ITT) analysis will be affected.
- Counteracting potential loss to follow-up is much easier as it will not affect ITT analyses results.

30

Varying the variability
(increasing the spread of values)

- 80% power, 2-sided p≤0.05, 100% adherence
- HbA1c difference 1.5%

Variability (SD)	Sample size
3%	
4%	
5%	
6%	

https://bit.ly/sampsize-calc

31

Varying the HbA1c difference
(even more reality)

- 80% power, 2-sided p≤0.05
- 100% adherence, SD=3%

Difference	Sample size
1.5%	
1.2%	
1.0%	
0.75%	

32

Real life

	Scenario	Sample size
1	1.5% diff., 80% power, p≤0.05, SD=3%, 100% adherence	126
2	1.0% diff., 80% power, p≤0.05, SD=5%, 80% adherence	
3	1.0% diff., 90% power, p≤0.01, SD=5%, 80% adherence	

https://bit.ly/sampsize-calc

33

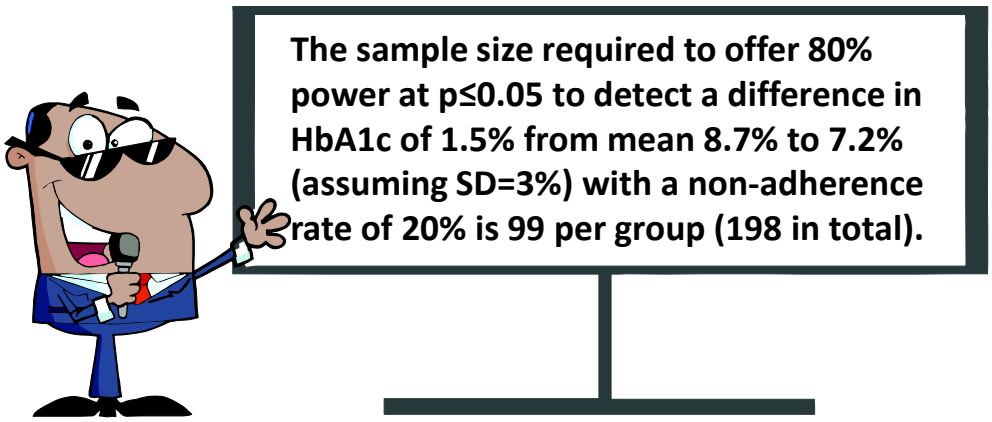
Sample size calculation statement

- Sample size justification.
- Method(s) of calculation(s).
- Quantities used in calculation(s):
  - Variance/standard deviation,
  - Mean values,
  - Response rate(s),
  - Difference(s) to detect.

34

Sample size calculation statement

- All the necessary details must be included for a reviewer / reader to be able to reproduce the calculations:



35

Example 2

- RCT is to compare the rate of gastrointestinal bleeding in patients taking Drug A vs. Drug B over 6 months.
- It is expected that relative rate of bleeding will by decreased by 50% (from 10% to 5%).

36

https://bit.ly/sampsize-calc

Rate of event in NON-TREATED/placebo (group 1) (%)	10	INPUT
Rate of event in TREATED/experimental (group 2) (%)	5	RESULTS
alpha	0.05	type 1 error rate
Power (%)	80	1-beta
1- or 2 tail	2	
group 1 to group 2 ratio	1	ignored if paired design!
Study design	unpaired	
Drop-outs (%)	0	
Drop-ins (%)	0	
Loss to follow up (%)	0	
n group 1	432	
n group 2	432	
n (per study)	864	
Relative risk of event in treated group vs. placebo (%)	50.00	i.e. 50% risk reduction
Odds ratio treated group vs. placebo	0.47	
Observed number of events	112	
Number of events in placebo group under H0	75	
Number of events in experimental group under H0	37	

Questions/comments: Andrzej.Januszewski@sydney.edu.au

37

Relative risk vs. Odds ratio

- **Relative Risk (RR):** "How many times more likely?"
- Definition: The ratio of the probability of an event in the exposed group to the probability in the unexposed group.
- Interpretation: “People taking Drug A are X times/% as likely to get better* compared to those not taking it.”

- **Odds Ratio (OR):** "How many times greater are the odds?"
- Definition: The ratio of the odds of the event in the exposed group to the odds in the unexposed group.
- Interpretation: “The odds of getting better* on Drug A are X times the odds without it.”

38

Example 3: Surveys / questionnaires

- **Example Project:** Evaluating Student Satisfaction with Online Learning.
- **Research Question:** "How satisfied are university students with online learning compared to in-person classes after the COVID-19 pandemic?"
- Study Design Overview
  - **Type:** Cross-sectional survey;
  - **Participants:** Undergraduate students enrolled at a single university;
  - **Total population (N):** 2,000 enrolled undergrads;
  - **Method:** Online questionnaire sent by email;
  - **Tool:** A custom questionnaire with Likert-scale questions (e.g. 1–5 satisfaction levels), plus demographics.

39

https://bit.ly/sampsize-calc

BEFORE COMMENCING THE SURVEY			**INPUT**
Confidence level (%)	95	How reliable we want the results to be?	**RESULTS**
Margin of error (e) (%)	5.00	How much the opinions and behaviours of the population you survey is likely to deviate from the total population?	
Total population available (N)	2000		
Response distribution (p) (%)	50	Percentage of your surveyed population that picks a particular answer. 50% is a conservative	
Response rate (%)	25	Percentage of questionnaires returned (1-100%)	
Sample size required	**1292**	Taking the surveyed population size and response rate into consideration	
AFTER COLLECTING THE SURVEY			
Number or returned questionnaires	323	How many questionnaires were actually	
Margin of error (e) achieved (%)	**4.994**		

Questions/comments: Andrzej.Januszewski@sydney.edu.au

40

Summary

- Sample size considerations are important
  - Influenced by desired power, treatment differences, variability and adherence
- Sample size estimation does not need to be very accurate, only adequate!
- Do not be afraid to seek help if you find the process too complicated!

41

Sample size needs to be just right…

- Too small  = weak study
- Too big  = wasteful
- Just right  = efficient and ethical

https://bit.ly/sampsize-calc
https://bit.ly/sampsize-formula

42