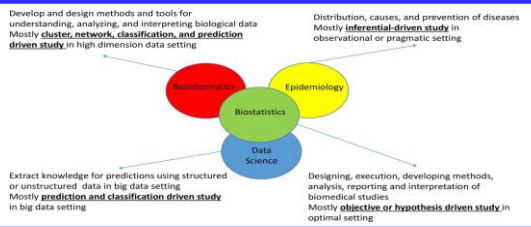
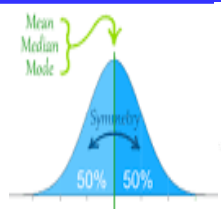
Using Statistics for Research and Clinical Practice

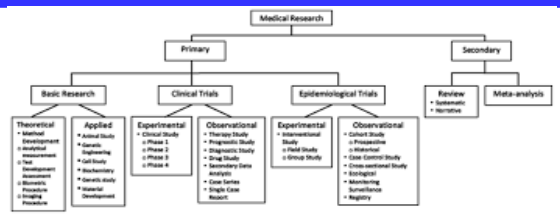
Sadik A. Khuder, DDS, Ph.D.,

College of Medicine

University of Toledo







Why you need to use Statistics?

- **The use of statistics in bio-medical journals has increased dramatically over the past few decades.**
- **Practitioners need to understand statistics well enough to follow and evaluate the empirical studies that provide an evidence base for their practices.**
- **Clinicians practice with individual patients, while conclusions about care practices almost always involve considerations of aspects of the clinical courses followed by many.**
- **Statistics is one of the important tools to help bridge this gap.**

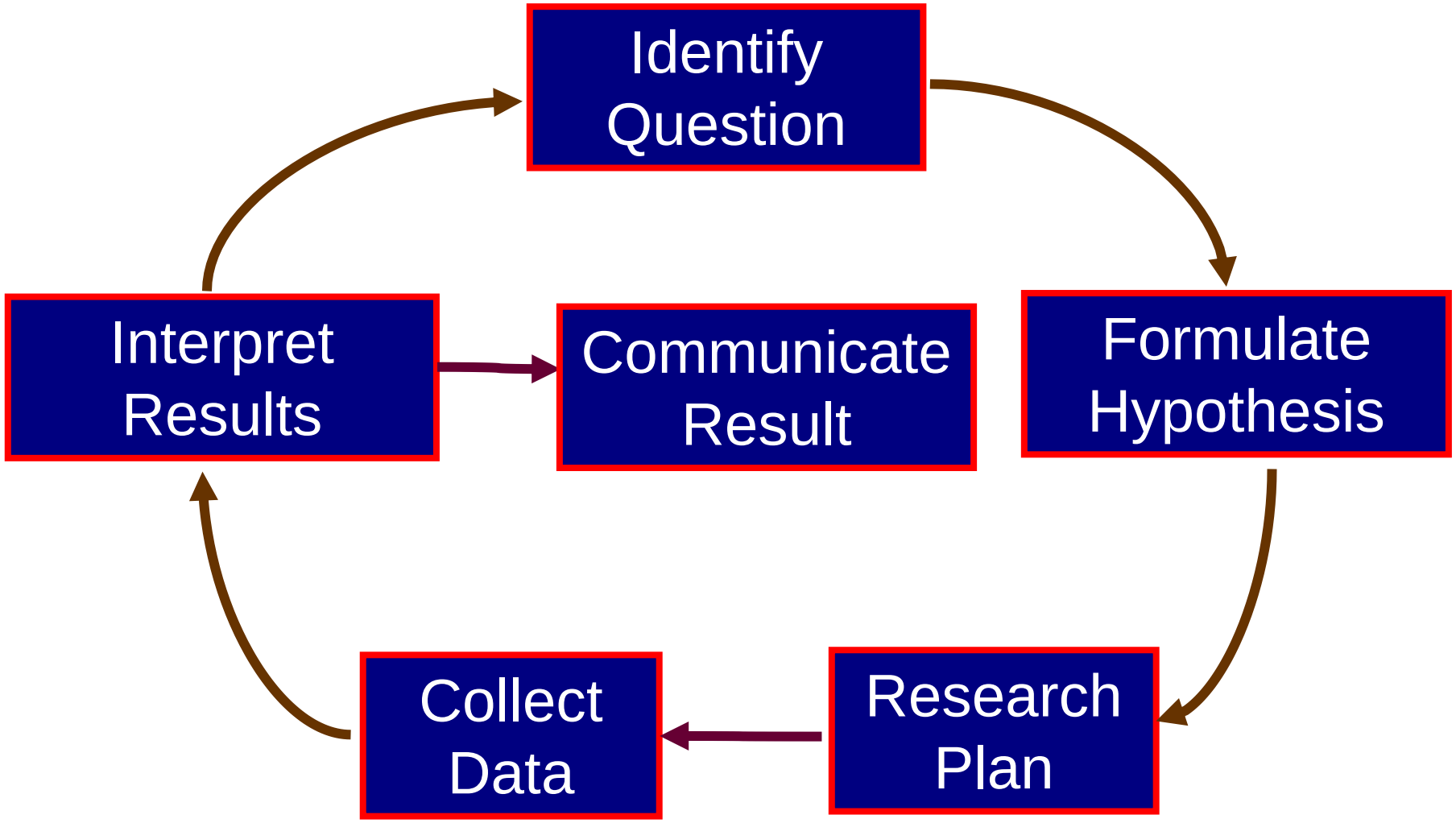
Statistical analysis

- Statistical analysis is a crucial part of a research.
- A scientific study must include statistical tools in the study, beginning from the planning stage.
- Statistical methods provide a way for formally accounting for sources of variability in the study.
- The use of statistics allows the researcher to form reasonable and accurate inferences from collected information, and make sound decisions in the presence of uncertainty.
- Statistics are key to preventing errors and biases in biomedical research.

Statistics for Research

- In biomedical research, sound statistics is essential for interpreting and reproducing results and thus avoiding the unnecessary and unethical use of subjects.
- Mistakes in the experimental design, statistical analysis, interpretation of p-values, and presentation of the findings, can result in:
  - Ethical and financial costs
  - Low success rates of subsequent clinical trials or technology development.

The Research Method



Statistical Errors in Scientific Studies

1) Flawed and inadequate hypothesis
2) Improper study design
   A. Inadequate sample size
   B. Lack of adequate control condition/group
3) Overstatement of the analysis results
   A. Excessive interpretation of limited or insignificant results
   B. Confusion between P value and clinical significance
   C. p-hacking
   D. Confusion of correlations, relationships, and causations
4) Inappropriate presentation of the results and effects

What is a Hypothesis?

- A statement about a specific research question, and it outlines the expected result of the experiment.
- Hypotheses are sometimes called “educated guesses”, but they are in fact based on previous observations, existing theories, scientific evidence, and logic.
- A study is only as good as its hypothesis
- The two hallmarks of a scientific hypothesis are
  1) Falsifiability
  2) Testability

Falsifiability

"No amount of experimentation can ever prove me right; **a single experiment** ***can prove me wrong."***

Albert Einstein

- We can falsify statements, but we can not prove them.

Proving a hypothesis

- Someone claims that all swans are white.

- Confirmatory evidence cannot prove the assertion to be true.

- Contradictory evidence makes it clear the claim is invalid.

Hypotheses

- In patients with acute myocardial infarction (AMI), does the administration of intravenous nitrate (IN), as compared with none, reduce mortality?

- The null hypothesis (H_0) would be that administration of IN has no effect on mortality rate (MR) in AMI patients.

- The alternative hypothesis (H_1) would be that administration of IN decreases MR in AMI patients.

H_0: $MR_{IN} = MR_{none}$

H_1: $MR_{IN} < MR_{none}$

What is a “Proper” Hypothesis?

- **A clear, testable statements written in the present tense that includes practical reasoning**
- **To begin formulating a hypothesis:**

1) **Review all the information gathered during research**
2) **Figure out what the main question of the study is**
3) **Form a general statement outlining this question and the overall expectation of the experiment**

The "PICOT" Model

Example: **Patients using cholesterol-lowering drug A for 6 months will have lower cholesterol level than those using drug B.**

Population- the specific group or individual the research pertains to (Patients with high cholesterol level)

Interest- the main concern of the study (Effects of drug on cholesterol level)

Comparison- the main alternative group (Drug A vs Drug B)

Outcome- what result is expected (Lower cholesterol level)

Time- the length of the experiment (6 months)

Bad Hypothesis Examples

Bad hypothesis	Prediction/research question	Problem
Garlic prevents smallpox.	**Participants who eat garlic daily will not be affected by smallpox.**	**Nobody gets affected by smallpox—*not falsifiable*.**
Drug A is better than drug B.	***??***	**No clearly defined variables - *not testable*.**

Statistical Test

- The goal of the test is to reject H_0 in favour of H_1.

Statistical decision	True state of H_0	
	H_0 is false	H_0 is true
Reject H_0	Correct	Type I error α
Do not reject H_0	Type II error β	Correct

Types of Errors

Type I Error
(false positive)

This kid has Dry Mouth

Type II Error
(false negative)

This kid has No Dry Mouth

Statistical Errors in Scientific Studies

1) Flawed and inadequate hypothesis
2) Improper study design
   A. Inadequate sample size
   B. Lack of adequate control condition/group
3) Overstatement of the analysis results
   A. Excessive interpretation of limited or insignificant results
   B. Confusion between P value and clinical significance
   C. p-hacking
   D. Confusion of correlations, relationships, and causations
4) Inappropriate presentation of the results and effects

Improper study design

100% of all disasters are failures of design, not analysis.

-- Ron Marks, Toronto, August 16, 1994

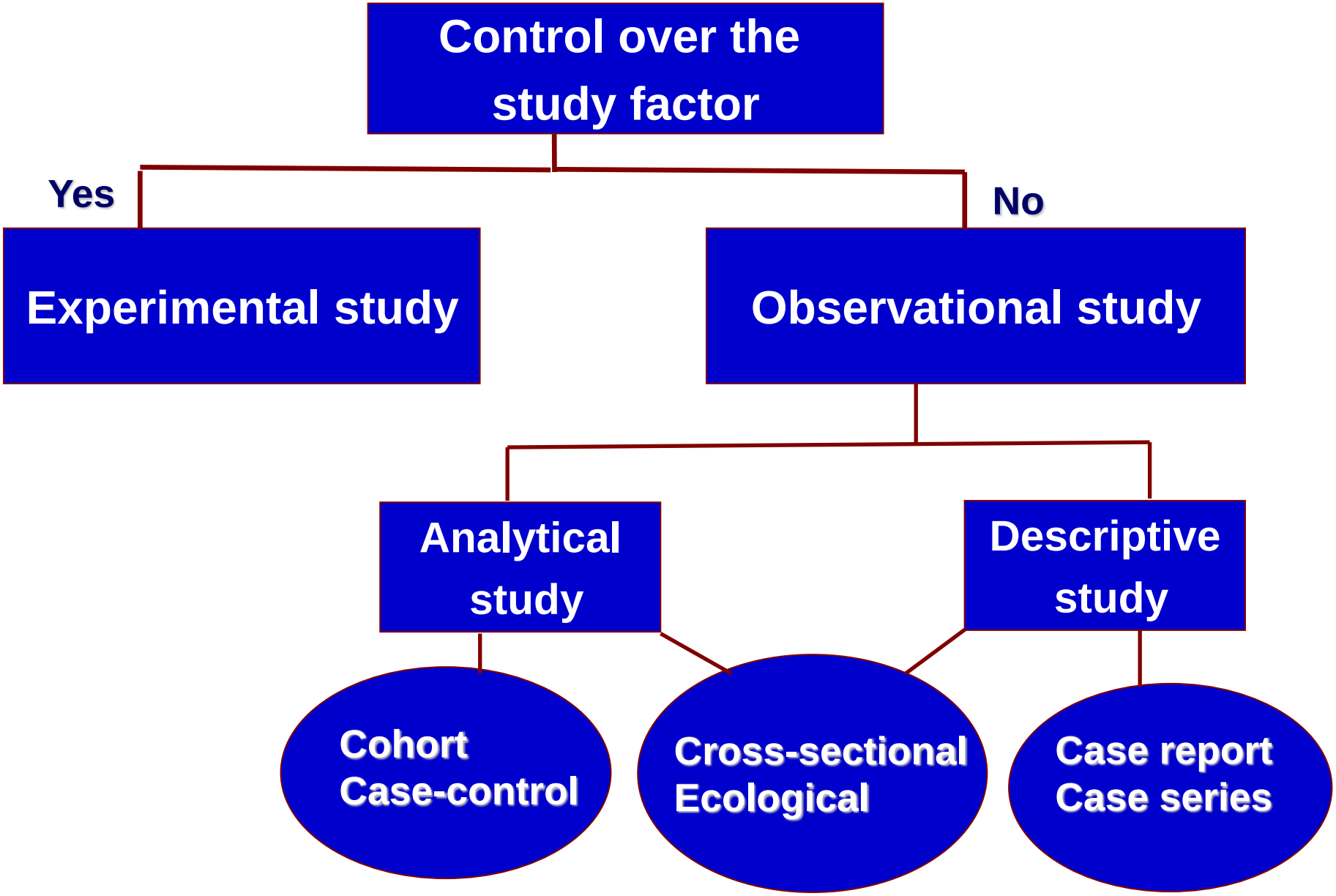
To propose that poor design can be corrected by subtle analysis techniques is contrary to good scientific thinking.

--Stuart Pocock (Controlled Clinical Trials, p 58)

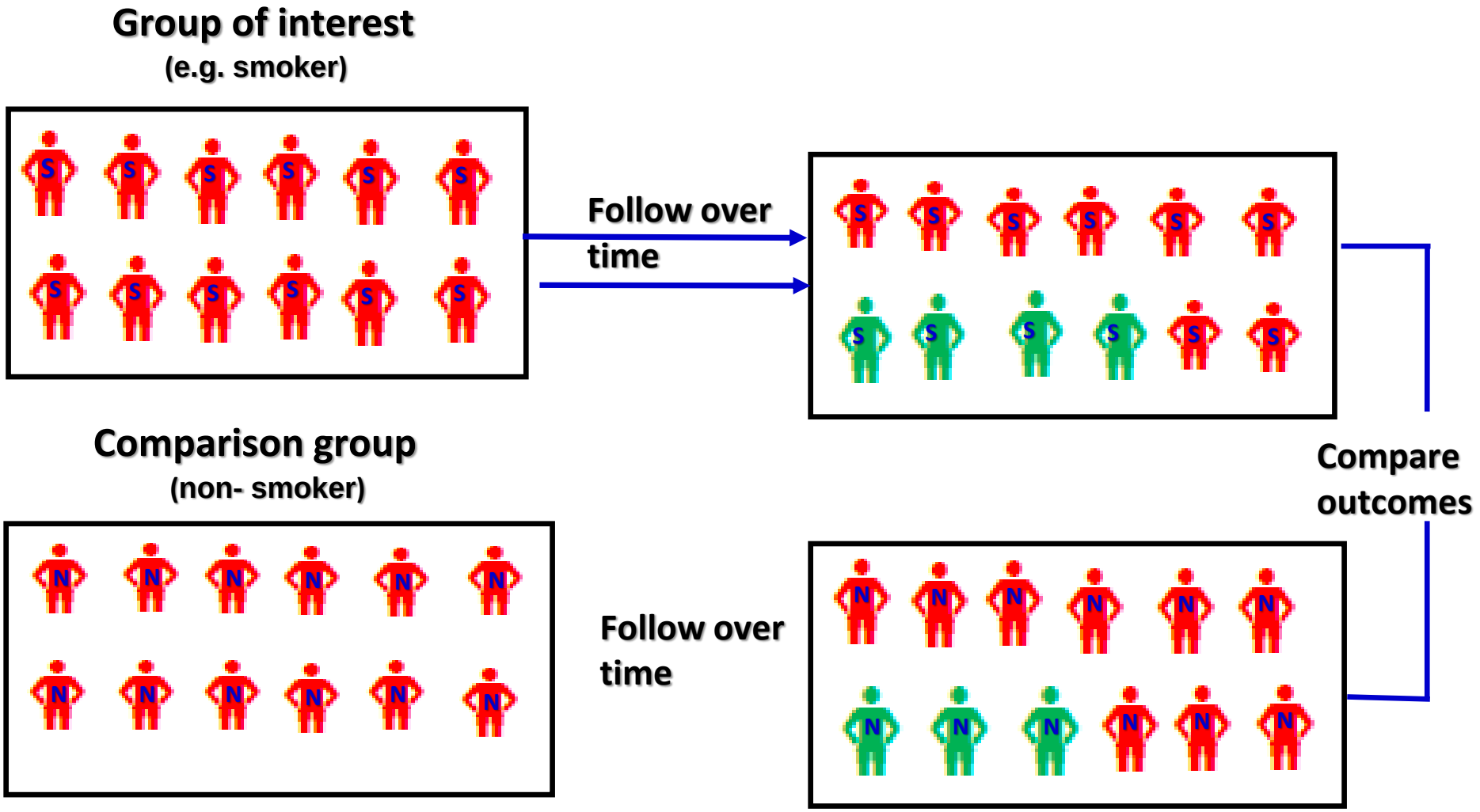
Study Design

- Designs are arrangements/patterns for obtaining/producing data
- A design must address the following issues:
  - How many subjects to include?
  - How to select the subjects?
  - How to form groups if needed?
  - What variables to measure?

Anatomy of Research Studies



Cohort Studies

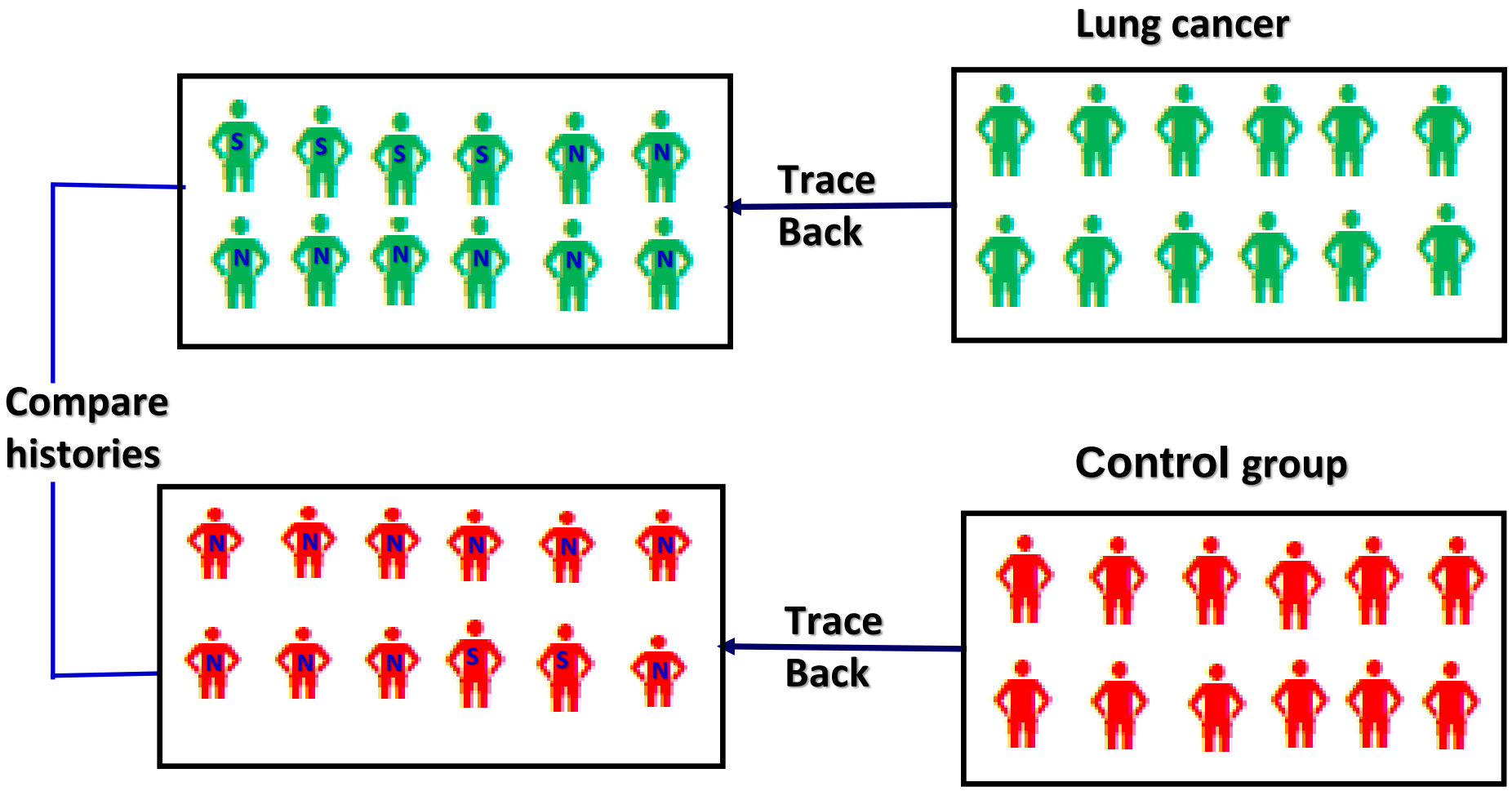


Risk among smoker = 4/12 = 0.3

Relative Risk = 0.33 / 0.25 = 1.3

Risk among non-smoker = 3/12 = 0.25

Case-control Studies

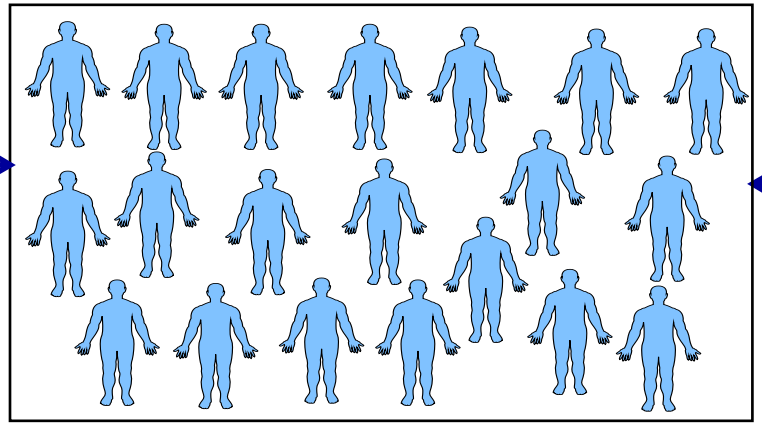


Odds of smoking in cases = 4/8 Odds of smoking in control = 2/10

Odds ratio = 0.5/0.2 = 2.5

Cross-sectional Studies

Take histories



Assess outcomes

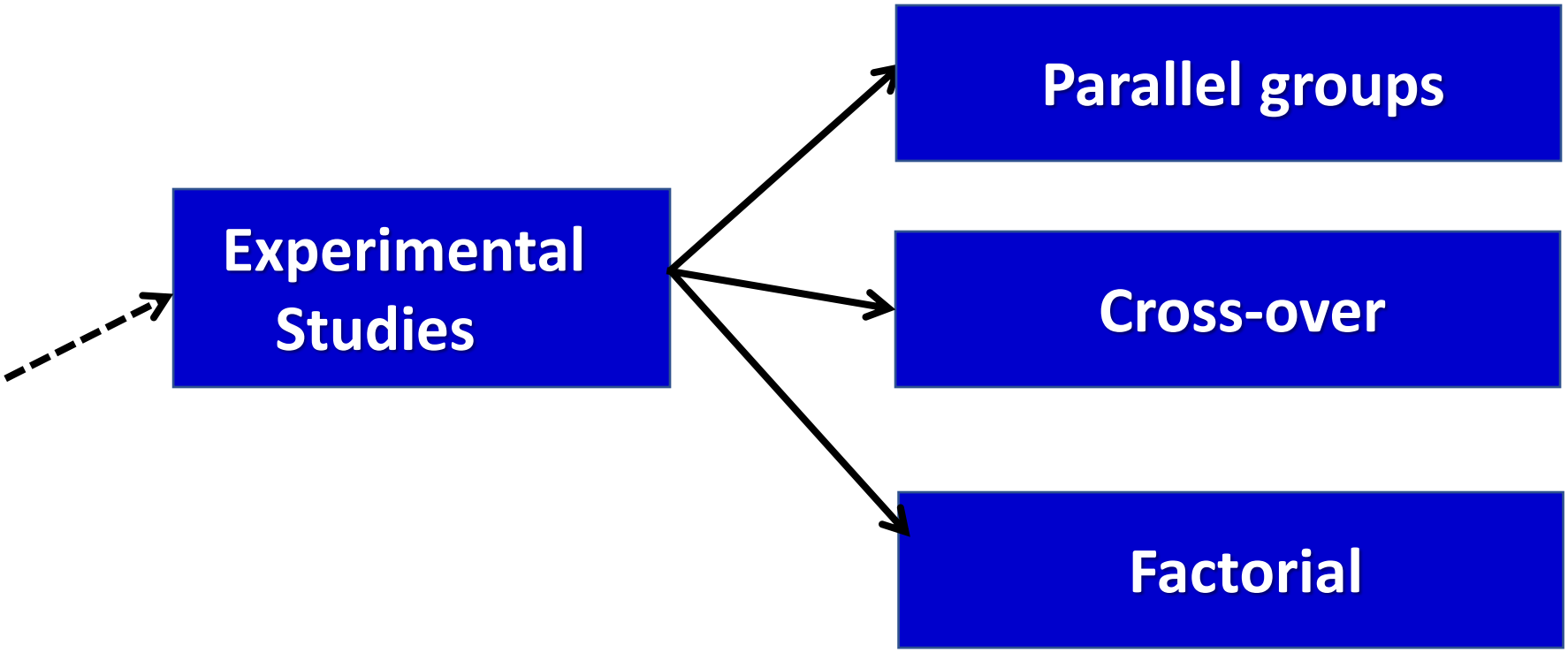
Odds of smoking in cases = 4/5

Odds of smoking in control = 4/11

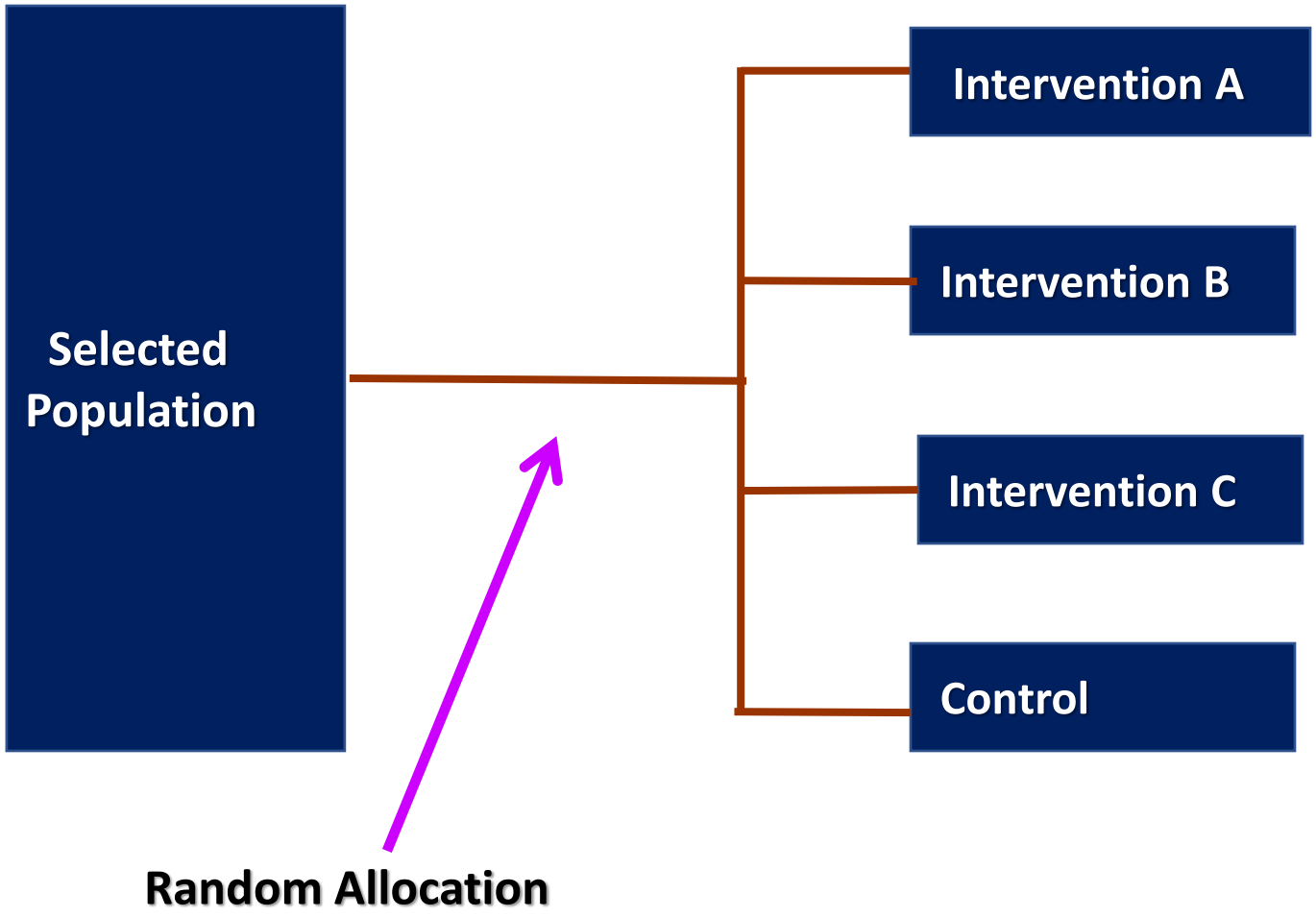
Odds ratio = 0.8/0.36 = 2.2

Relate histories to outcomes

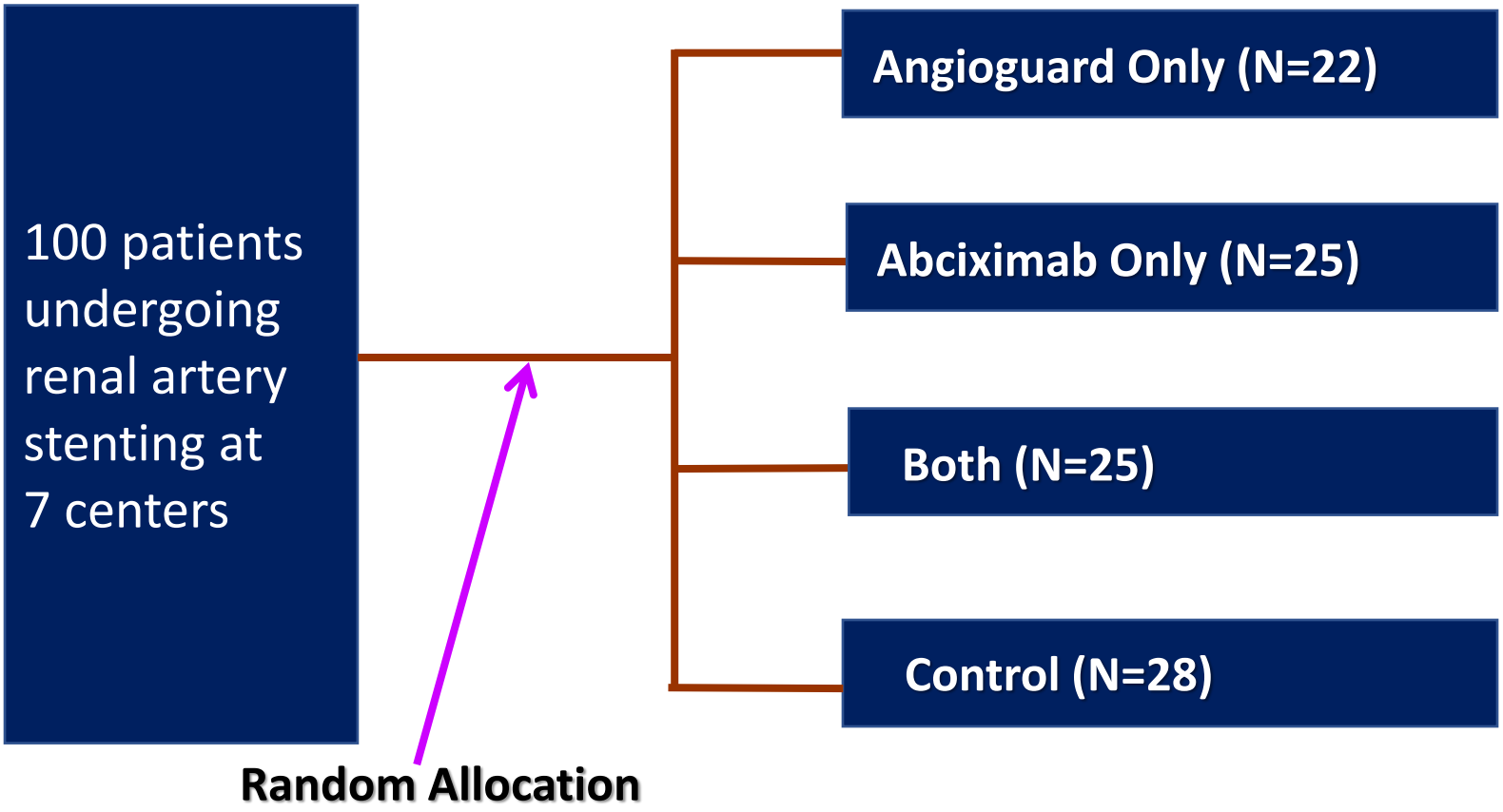
Types of study designs



Parallel Design

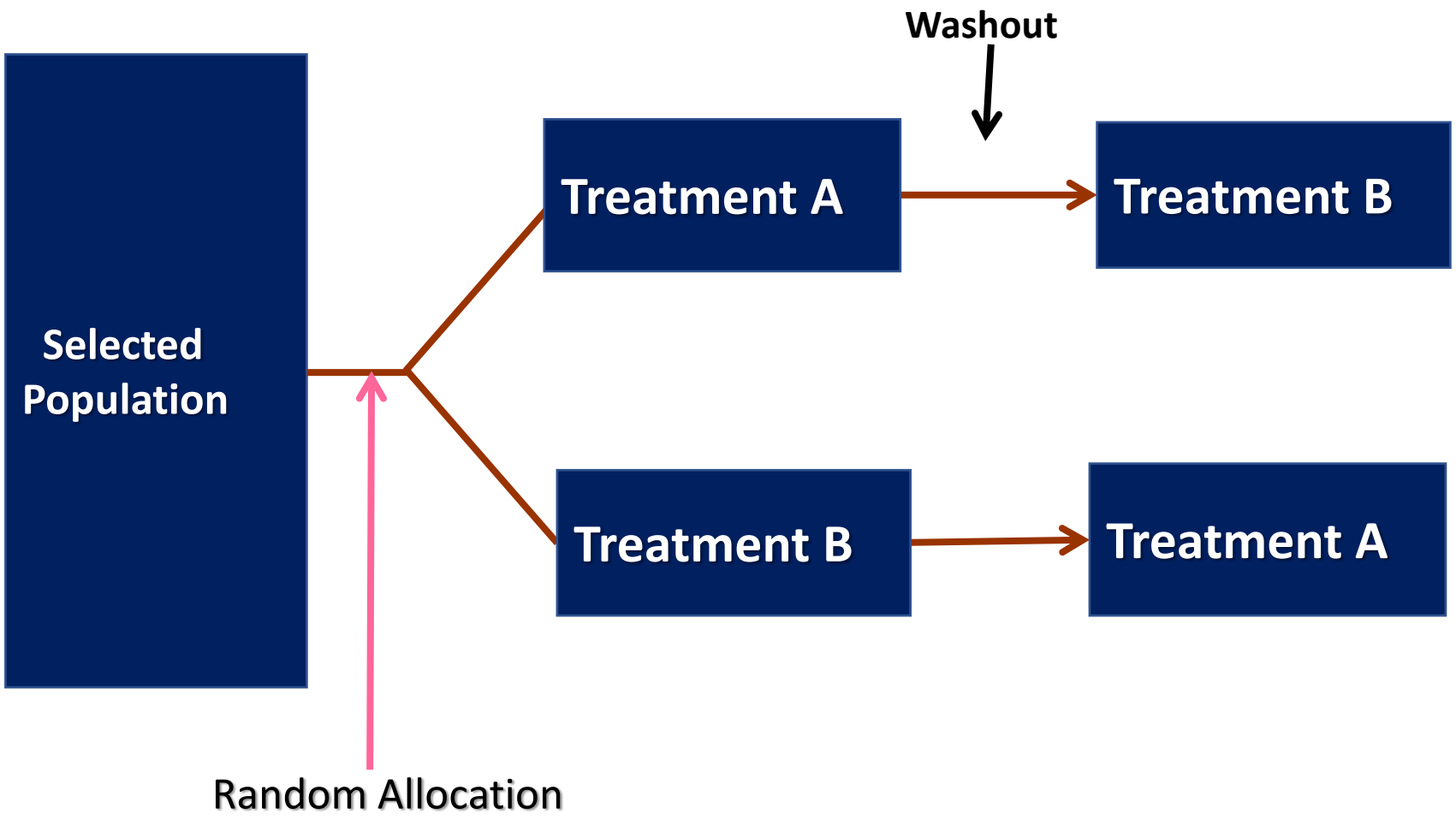


Factorial Design



Cooper et al. Circulation 2008 May 27;117(21):2752-60

Crossover Design



Type of study most appropriate with each objective

Table 1

Objective	Common design
Prevalence	Cross sectional
Incidence	Cohort
Cause (in order of reliability)	Cohort, case-control, cross sectional
Prognosis	Cohort
Treatment effect	Controlled trial

How many subjects to include?

- We can draw a precise and accurate conclusion only with an appropriate sample size.
- A smaller sample will give a result which may not be sufficiently powered to detect a difference between the groups and the study may turn out to be falsely negative leading to a type II error.
- Very often, a small sample size is decided arbitrarily based on the researchers' convenience, available time, and resources, resulting in a null trial due to insufficient number of subjects studied.
- In a JAMA study, researchers found that out of 102 null trials, only 36% had 80% power to detect a relative difference of 50% between groups.

JAMA. 1994;272:122–124.

How many subjects to include?

- **A very large sample size is also not recommended as it has its own consequences.**
  1) **It is a waste of the limited available resources in terms of time and money when an answer can be accurately found from a smaller sample.**
  2) **Recruiting more subjects than required can also be termed as "unethical" as the patients participate in a study with faith and an altruistic motive which should not be mis utilized.**
  3) **In randomized controlled trials more people will be denied a better regimen and will get a placebo or an inferior treatment with its associated side effect or toxicity due to the inherent design of the study.**

Why to calculate sample size and power?

- To show that under certain conditions, the hypothesis test has a good chance of showing a desired difference (if it exists)
- To show to the funding agency that the study has a reasonable chance to obtain a conclusive result
- To show that the necessary resources (human, monetary, time) will be minimized and well utilized

Package 'RcmdrPlugin.EZR'

November 6, 2022

Type Package

Title R Commander Plug-in for the EZR (Easy R) Package

Version 1.61

Date 2022-11-11

Author Yoshinobu Kanda

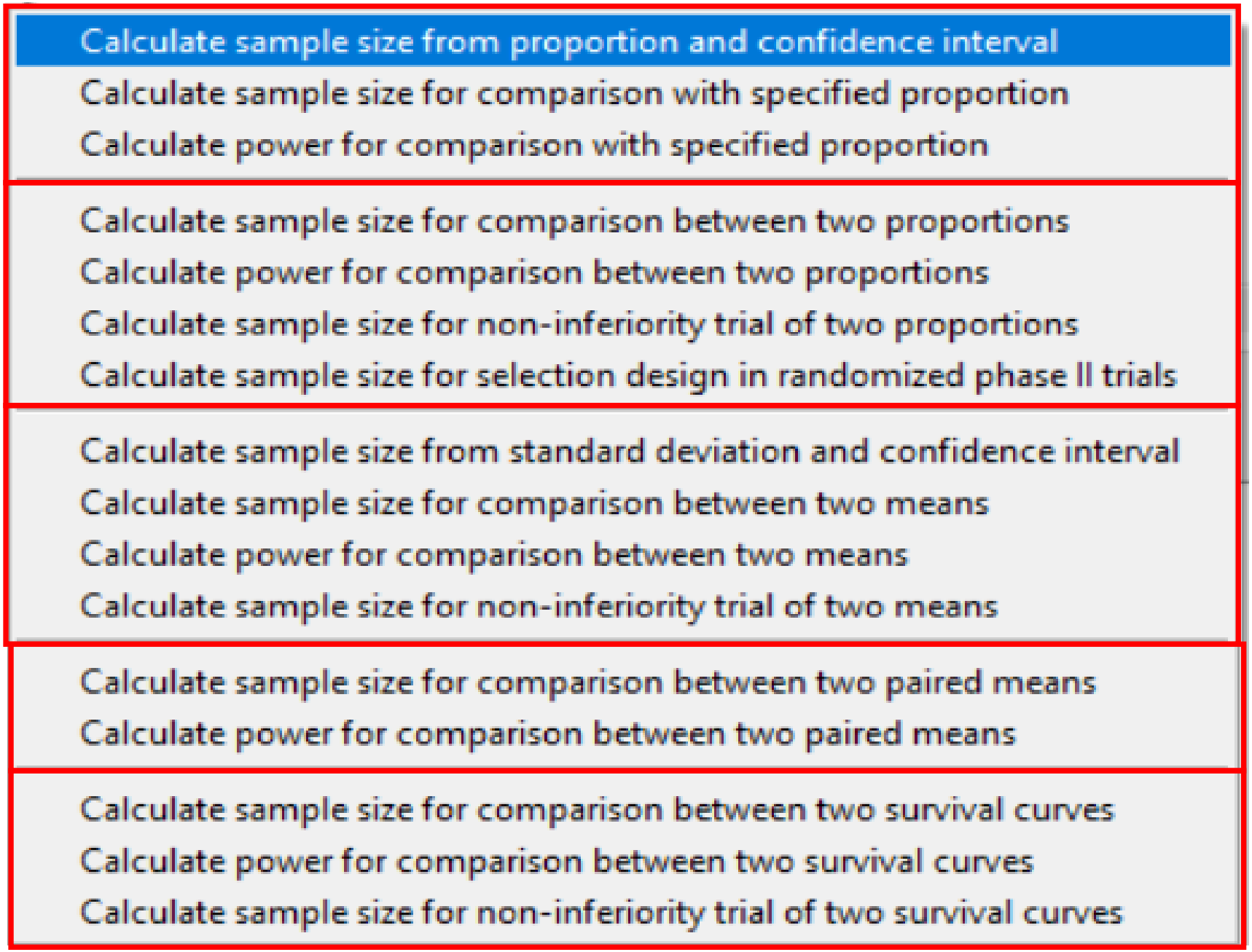
Maintainer Yoshinobu Kanda <ycanda-tky@umin.ac.jp>

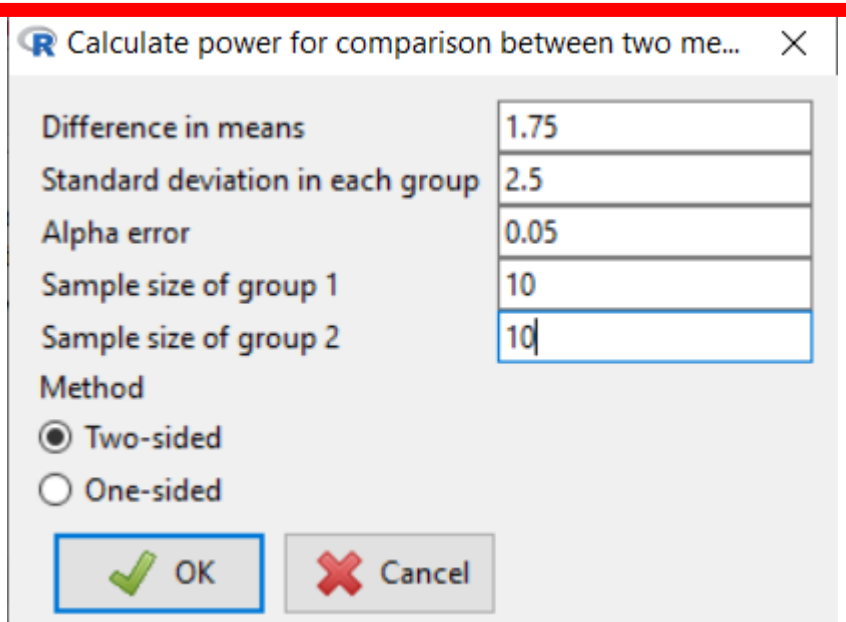
Depends R (>= 4.2.0)

Imports Rcmdr (>= 2.8.0), readstata13

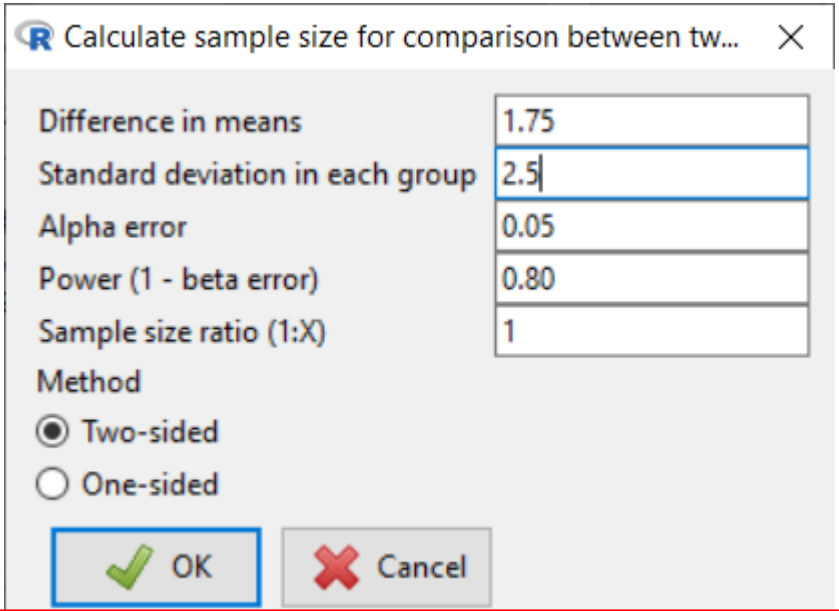
Suggests abind, aod, aplpack, brant, car, clinfun, cmprsk, foreign, ggplot2, lawstat, meta, metatest, netmeta, multcomp, mvtnorm, Matching, pROC (>= 1.15.0), survivalROC, survRM2, tableone, readxl, lmerTest, swimplot, currentSurvival

Description EZR (Easy R) adds a variety of statistical functions, including survival analyses, ROC analyses, metaanalyses, sample size calculation, and so on, to the R commander. EZR enables point-and-click easy access to statistical functions, especially for medical statistics. EZR is platform-independent and runs on Windows, Mac OS X, and UNIX. Its complete manual is available only in Japanese (Chugai Igakusha, ISBN: 978-4-498-10918-6, Nankodo, ISBN: 978-4-524-26158-1, Ohmsha, ISBN: 978-4-274-22632-8), but an report that introduced the investigation of EZR was published in Bone Marrow Transplantation (Nature Publishing Group) as an Open article. This report can be used as a simple manual. It can be freely downloaded from the journal website as shown below. This report has been cited in more than 3,000 scientific articles.

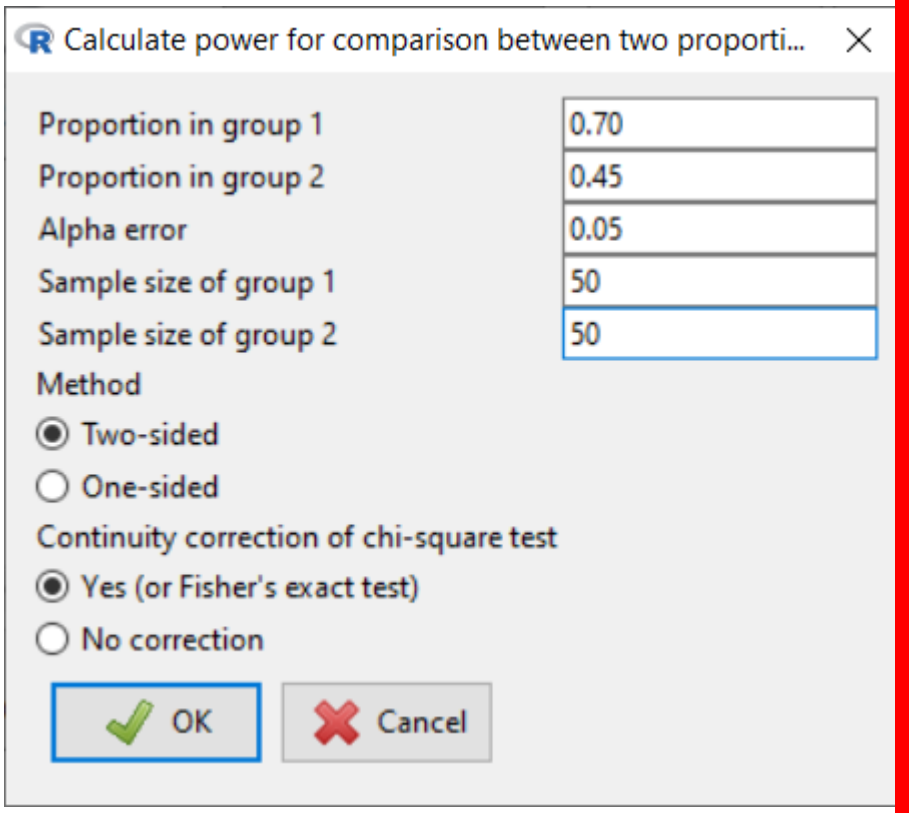

Calculate sample size from proportion and confidence interval
Calculate sample size for comparison with specified proportion
Calculate power for comparison with specified proportion
Calculate sample size for comparison between two proportions
Calculate power for comparison between two proportions
Calculate sample size for non-inferiority trial of two proportions
Calculate sample size for selection design in randomized phase II trials
Calculate sample size from standard deviation and confidence interval
Calculate sample size for comparison between two means
Calculate power for comparison between two means
Calculate sample size for non-inferiority trial of two means
Calculate sample size for comparison between two paired means
Calculate power for comparison between two paired means
Calculate sample size for comparison between two survival curves
Calculate power for comparison between two survival curves
Calculate sample size for non-inferiority trial of two survival curves



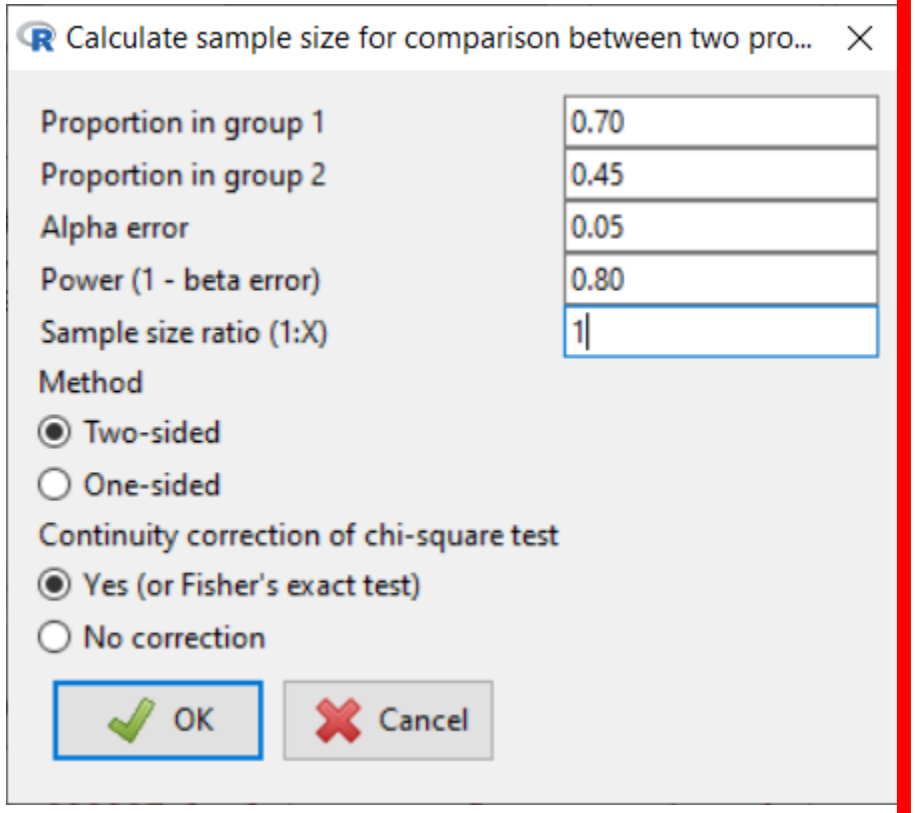

```
> PowerMean(1.75, 2.5, 0.05, 10, 2, 1)
                      Assumptions
Difference in means          1.75
Standard deviation            2.5
Alpha                        0.05
                         two-sided

Sample size
N1                             10
N2                             10

                         Estimated
Power                        0.347
```




```
> SampleMean(1.75, 2.5, 0.05, 0.80, 2, 1)
                      Assumptions
Difference in means          1.75
Standard deviation            2.5
Alpha                        0.05
                         two-sided
Power                         0.8
N2/N1                           1

Required sample size     Estimated
N1                              33
N2                              33
```




```
> PowerProportion(0.70, 0.45, 0.05, 50, 2, 1, 1)
            Assumptions
P1                  0.7
P2                 0.45
Alpha              0.05
               two-sided
Sample size
N1                   50
N2                   50

               Estimated
Power              0.644
```




```
> SampleProportion(0.70, 0.45, 0.05, 0.80, 2, 1, 1)
                        Assumptions
P1                              0.7
P2                             0.45
Alpha                          0.05
                           two-sided
Power                           0.8
N2/N1                             1

Required sample size      Estimated
N1                               69
N2                               69
```

Factors Affecting Sample Size

1) Size of the difference you want to detect – The smaller the size of the difference in the outcome of interest you want to detect, the larger the number of participants who will need to compare.
2) The expected event rate in the control group and the treatment group.
3) Accepted probability of a type I error – α
4) Accepted probability of a type II error – β
5) Power – the higher the degree of certainty we require that the result we observe is a true result, then the greater the number of participants needed.
6) Study design – Different trial designs require different sample sizes.
7) Loss to follow up – Sample size should be adjusted to account for anticipated loss to follow up.

Reducing sample size

- Reduce the number of treatment groups being compared.
- Find a more precise measurement (e.g., average time to effect rather than proportion sick).
- Decrease the variability in the measurements.
  1) Make subjects more homogeneous.
  2) Use stratification.
  3) Control for other variables (e.g., weight).
  4) Average multiple measurements on each subject.

2.4. Sample size

Based on a previous study performed by our team members (26), and to find a 20% difference between the percentage of patients in abstinence (partial or total) between EG (37%) and CG (20%), for an alpha error of 5%, and statistical power of 80%, the size would be 220 subjects (110/group). Since it is a cluster randomization system, we will consider the "design effect" and we will assume a loss rate of 5%. Estimates of the intra-cluster correlation coefficient (ICC) in ECC by clusters in PC show that they are generally less than 0.05 (27). This ICC translates, for a cluster size of 15, into a design effect corresponding to a factor of 1.7. Assuming this value, the size would be 394 subjects to recruit (197 in each group).

Frontiers in Medicine 2022 1008832

Statistical Errors in Scientific Studies

1) Flawed and inadequate hypothesis
2) Improper study design
   - A. Inadequate sample size
   - B. Lack of adequate control condition/group
3) Overstatement of the analysis results
   - A. Excessive interpretation of limited or insignificant results
   - B. Confusion between P value and clinical significance
   - C. p-hacking
   - D. Confusion of correlations, relationships, and causations
4) Inappropriate presentation of the results and effects

Lack of adequate control condition/group

- **A study cannot be justified ethically unless it is capable of producing scientifically reliable results.**
- **Scientifically invalid research is unethical in that it exposes research subjects to risk without any possible benefit.**
- **In clinical trials, regardless of how good the results are in the intervention group, they count only when compared to the other group.**

The following episode related by Dr. E. E. Peacock

One day when I was a junior medical student, a very important Boston surgeon visited the school and delivered a great treatise on large number of patients who had undergone successful operations for vascular reconstruction. At the end of the lecture, a young student at the back of the room timidly asked, "Do you have any controls?" Well, the great surgeon drew himself up to his full height, hit the desk, and said, "Do you mean did I not operate on half the patients?" The hall grew very quiet then. The voice at the back of the room very hesitantly replied, "Yes, that's what I had in mind." Then the visitor's first really came down as he thundered, "Of course not. That wood have doomed half of them to their death." God, it was quiet then, and one could scarcely hear the small voice ask, "Which half?"

Medical World News, September 1, 1972, p.45

Lack of adequate control condition/group

Treatment

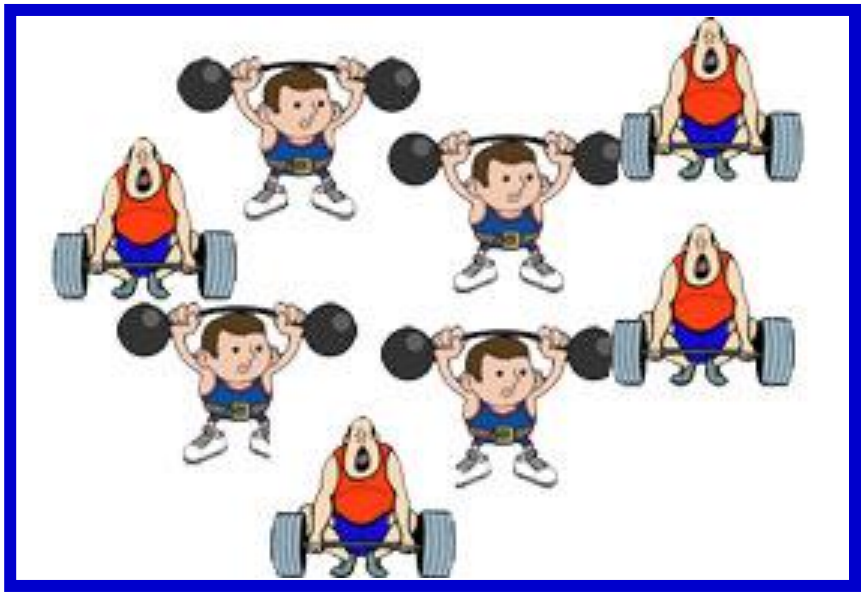
Control

- **To ensure that treatment group and Control group are equivalent at the beginning of the study, we can flip a coin for each person.**
- **That way each person has a 50% chance of being in either group – regardless of initial eating habits. This will then help us be sure that our results were a product of our treatment.**

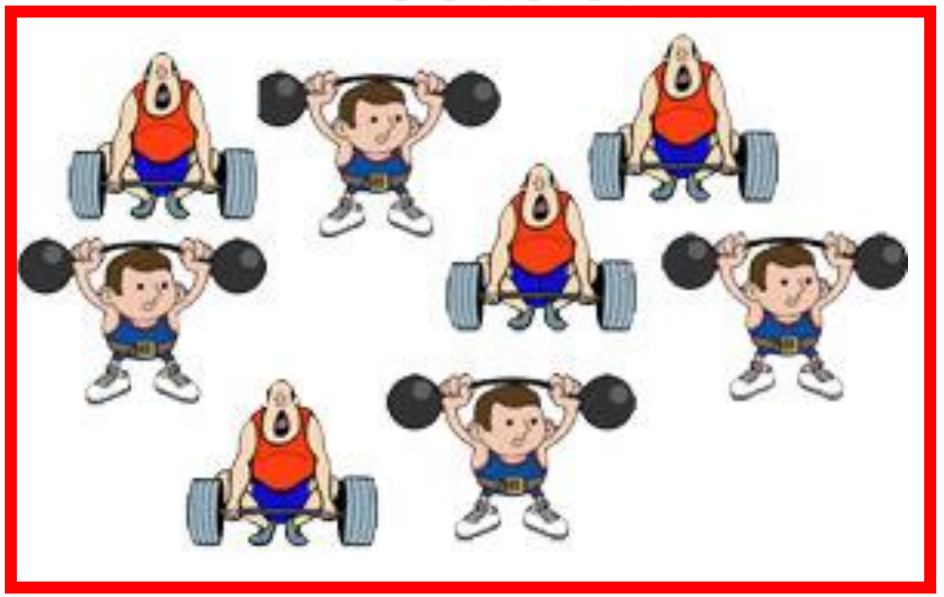
https://www.hydroassoc.org/research-101-randomization

Control Group

Treatment



Control

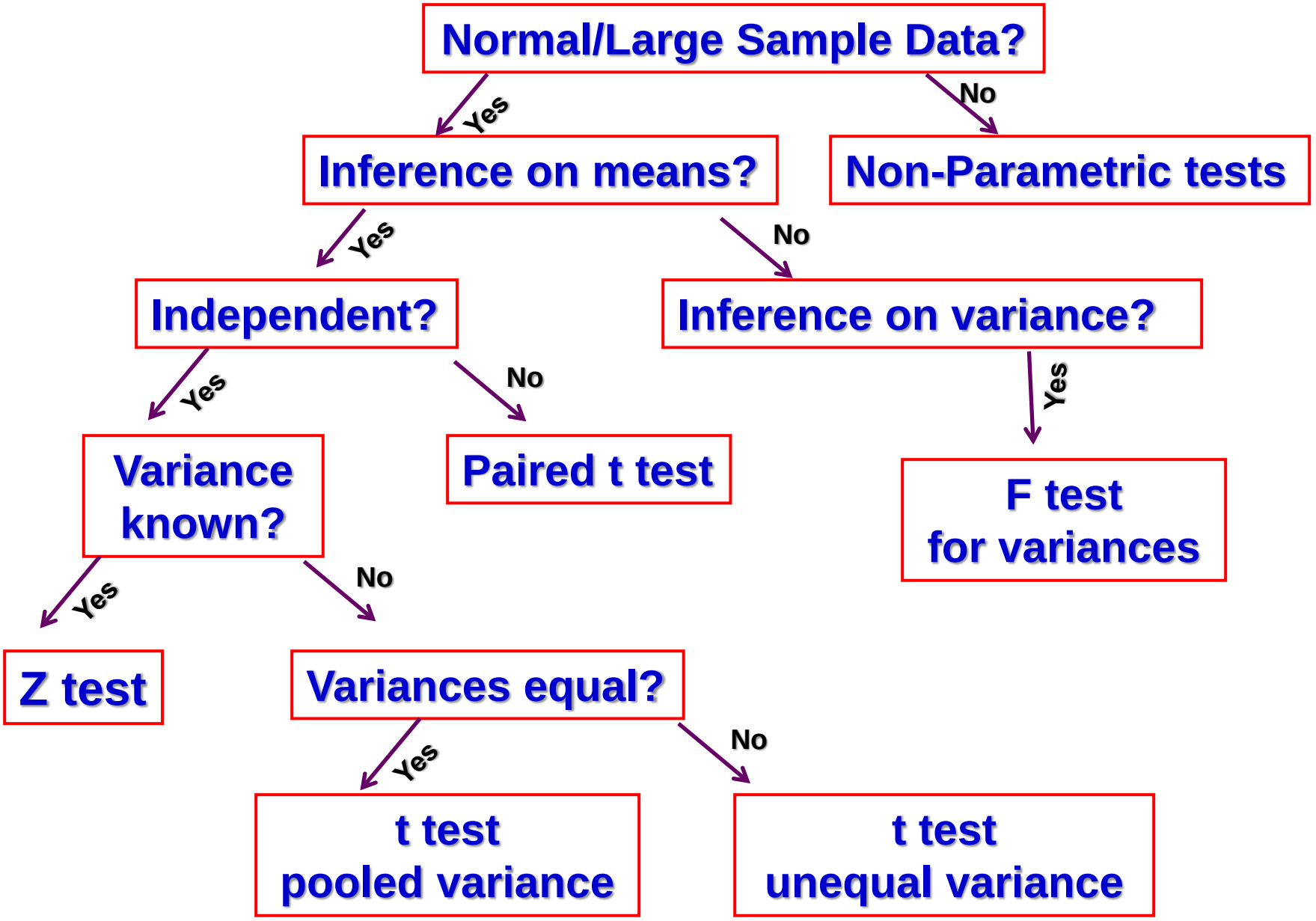


- **In this case randomization helped divide healthy and unhealthy people equally into treatment and control.**
- **Since Group A and Group B started off on equal footing, any difference in the outcome between the groups will be a result of the intervention.**

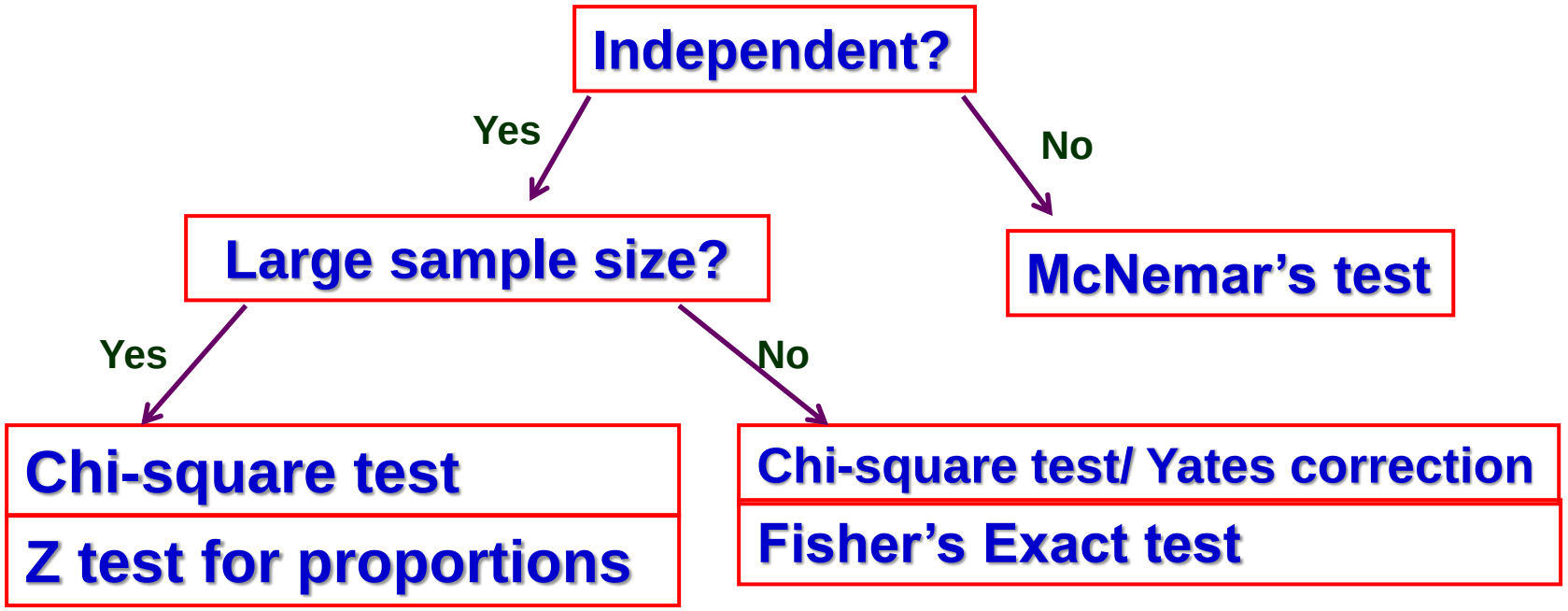
Statistical Errors in Scientific Studies

1) Flawed and inadequate hypothesis
2) Improper study design
   A. Inadequate sample size
   B. Lack of adequate control condition/group
3) Overstatement of the analysis results
   A. Excessive interpretation of limited or insignificant results
   B. Confusion between P value and clinical significance
   C. p-hacking
   D. Confusion of correlations, relationships, and causations
4) Inappropriate presentation of the results and effects

Statistical Tests for Continuous Data



Statistical Tests for Categorical Data



Common statistics for various types of outcome data

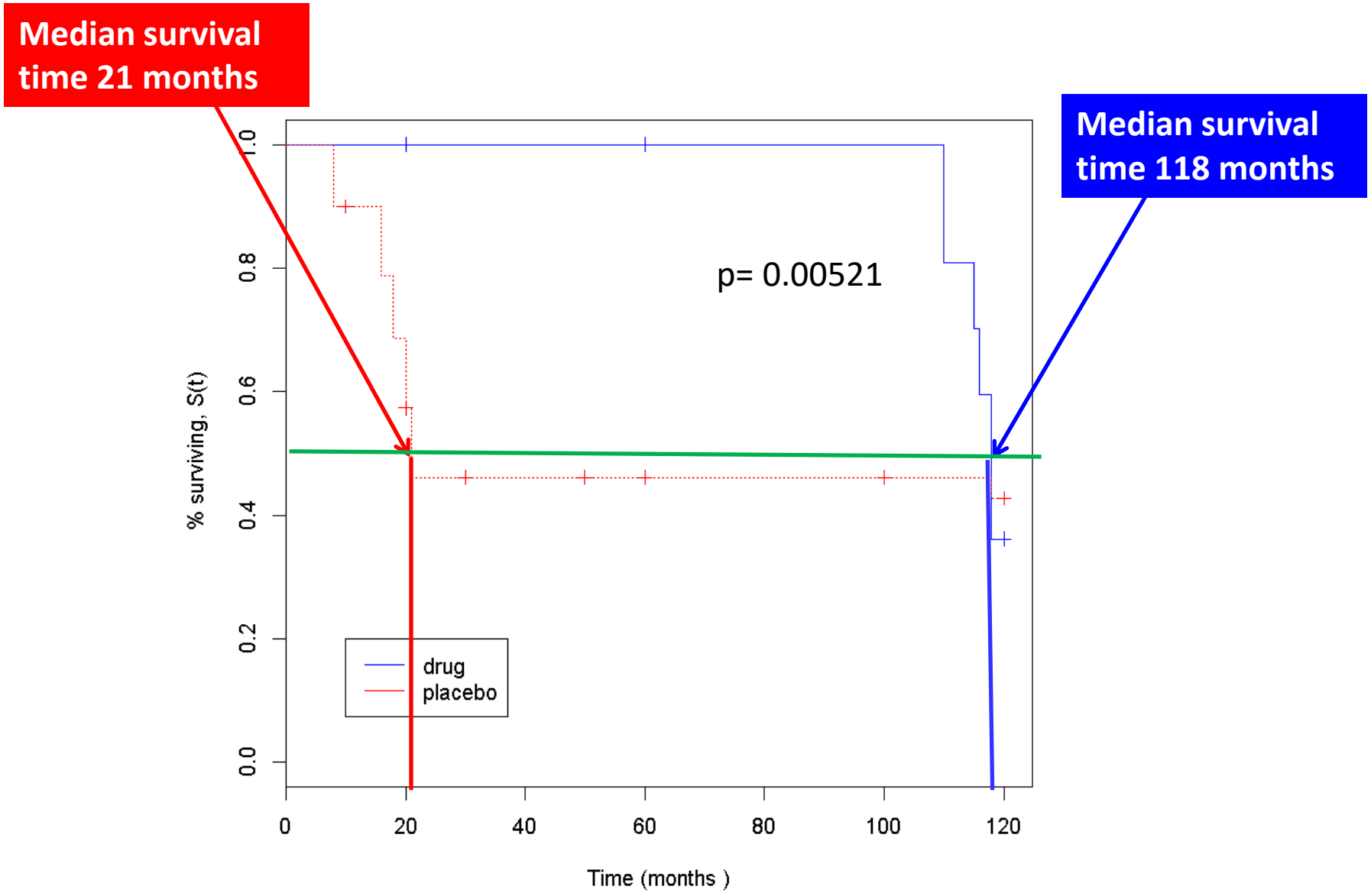
Outcome Variable	Are the observations independent or correlated?		Assumptions
	independent	correlated	
Continuous (e.g. pain scale, cognitive score)	t-test ANOVA Linear regression	Paired t-test RM ANOVA Mixed models	Outcome is normally distributed Outcome and predictor have a linear relationship
Binary or categorical (e.g. fracture yes/no)	Relative risks Chi-square test Logistic regression	McNemar's test Conditional logistic regression GEE modeling	Sufficient numbers in each cell (>=5)
Time-to-event (e.g. time to fracture)	Kaplan-Meier statistics Cox regression		Cox regression assumes proportional hazards between groups

Find a more precise measurement

Treatment	Mortality		Total
	Yes	No	
Drug	60	40	100
Placebo	50	50	100

```
        Pearson's Chi-squared test

data:  .Table
X-squared = 2.0202, df = 1, p-value = 0.1552
```

Find a more precise measurement



Use Correct Statistical Test

Cases	Matched Control	
	stored cooked food	did not store cooked food
stored cooked food	35	39
did not store cooked food	18	33

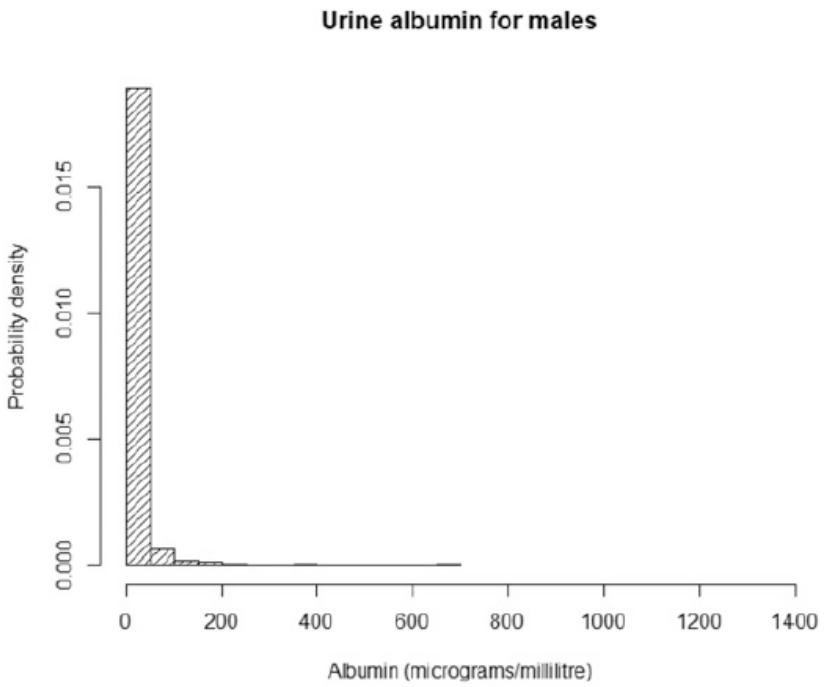
```
        Pearson's Chi-squared test with Yates' continuity correction

data:  t3
X-squared = 1.3236, df = 1, p-value = 0.25
```

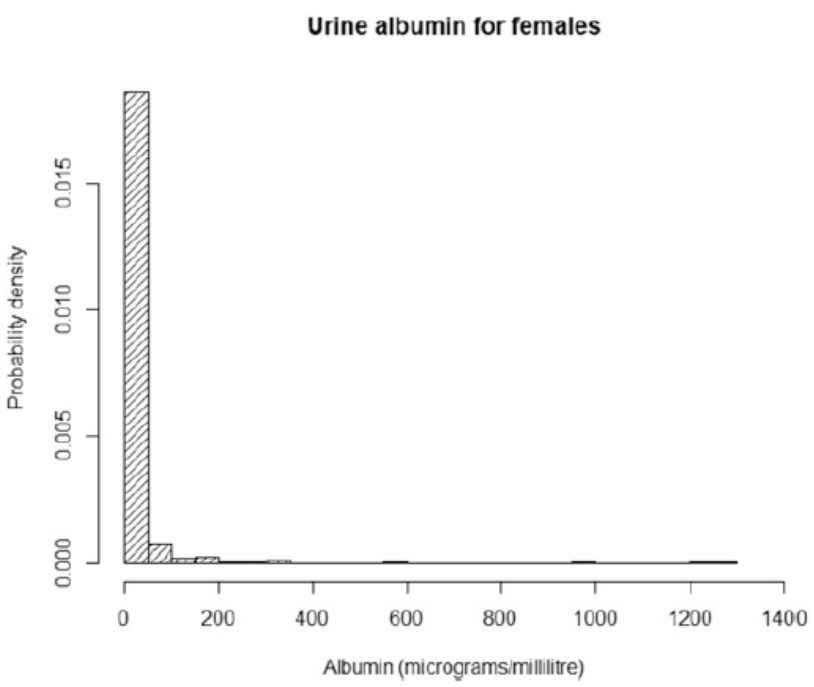
Use Correct Statistical Test

Cases	Matched Control	
	stored cooked food	did not store cooked food
stored cooked food	35	39
did not store cooked food	18	33

```
        McNemar's Chi-squared test with continuity correction

data:  t3
McNemar's chi-squared = 7.0175, df = 1, p-value = 0.008071
```

Best practice in statistics: The use of log transformation

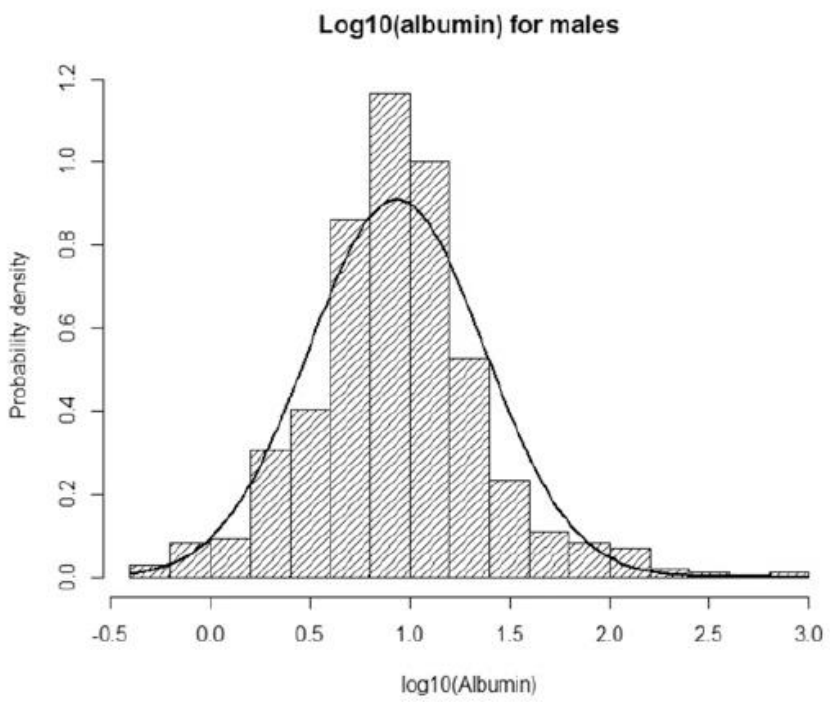


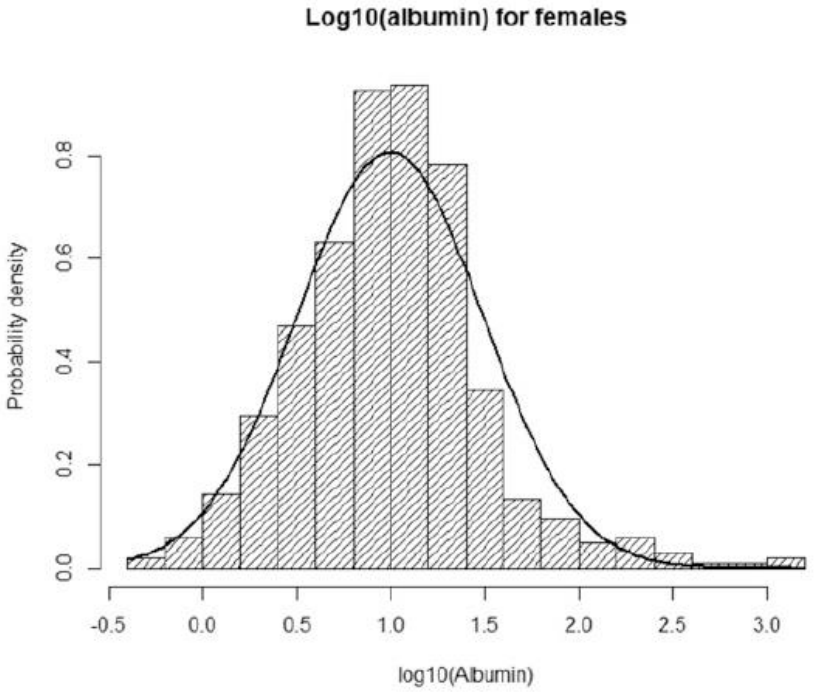


- Suitable summary statistics are the median and interquartile range (IQR).
- For men, median is 10.30 µg/mL, with an IQR of 4.55–15.47 µg/mL
- For women, median is 9.10 µg/mL, with IQR 5.35–19.30 µg/mL
- The means and standard deviations are not useful since the distribution is far from normal.

Annals of Clinical Biochemistry 2022, Vol. 59(3) 162–165

Best practice in statistics: The use of log transformation

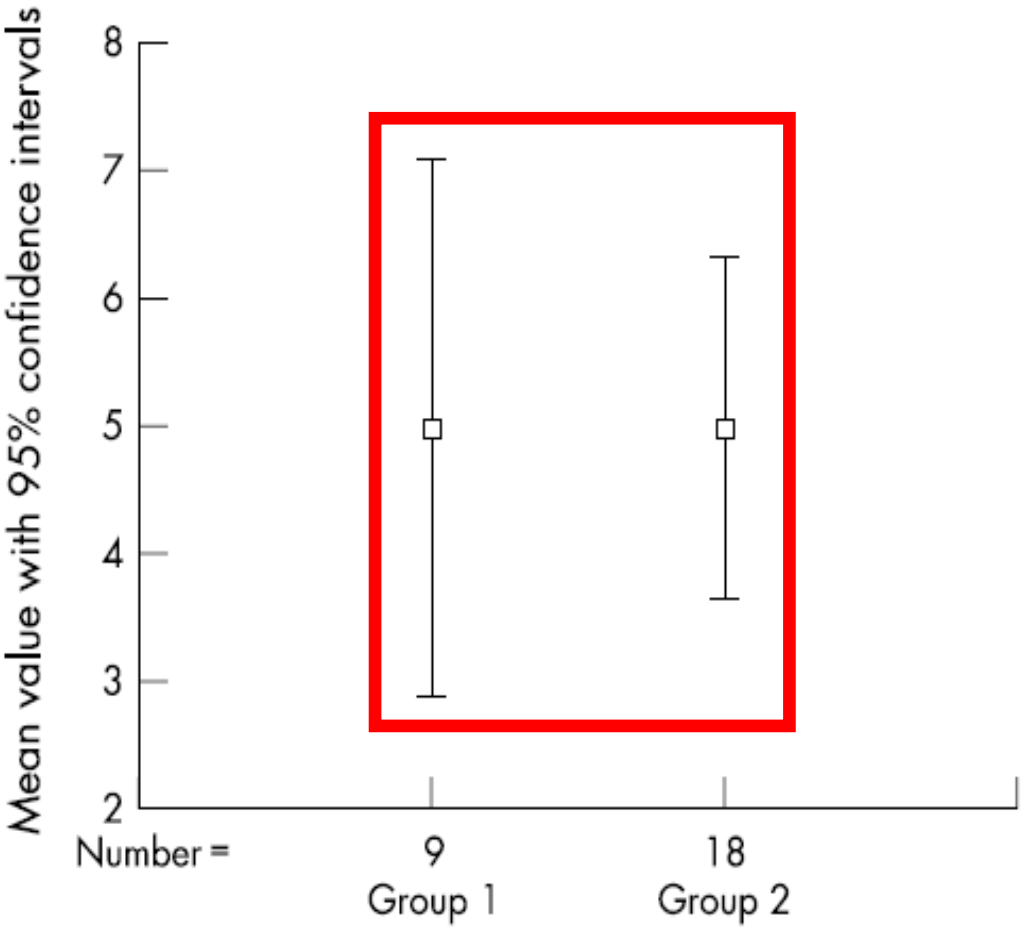




- The test statistic is t = 2.087 with 1012.2 degrees of freedom so that the p value is 0.0372
- According to this statistical test, there is a significant difference in the mean values of log10 (albumin) (0.933 for males and 0.995 for females).

Annals of Clinical Biochemistry 2022, Vol. 59(3) 162–165

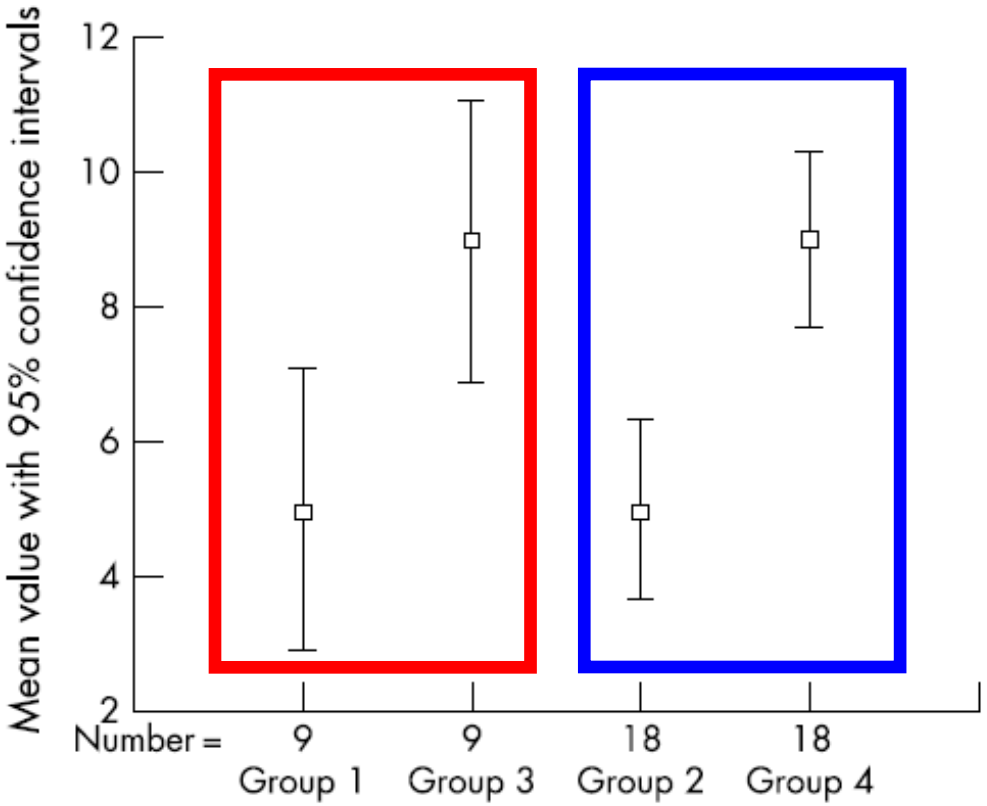
How increasing the number of subjects can give a more precise estimate of differences.



Group 1	Group 2
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
	1
	2
	3
	4
	5
	6
	7
	8
	9

Figure 1 Effect Change in confidence interval width with increasing numbers of subjects

How increasing the number of subjects can give a more precise estimate of differences.



Group 1	Group 2	Group 3	Group 4
1	1	5	5
2	2	6	6
3	3	7	7
4	4	8	8
5	5	9	9
6	6	10	10
7	7	11	11
8	8	12	12
9	9	13	13
	1		5
	2		6
	3		7
	4		8
	5		9
	6		10
	7		11
	8		12
	9		13

Figure 2 Effect of confidence interval reduction to demonstrate a true difference in means. This example shows that the initial comparison between groups 1 and 3 showed no statistical difference as the confidence intervals overlapped. In groups 3 and 4 the number of patients is doubled (although the mean remains the same). We see that the confidence intervals no longer overlap indicating that the difference in means is unlikely to have occurred by chance.

Clinical Significance vs Clinical Significance

- A P-value from statistical tests can only determine if there are differences between the two groups.
- It does not tell you whether one treatment group was better or worse than another group, or if the differences are actually clinically relevant.
- Just because something is statistically significant does not necessarily mean it's clinically important.
- Clinical significance measures the extent that a change can create a meaningful response for the patient.
  - For example, we determined that a new mouth wash formulation improved comfort in dry mouth patients by 1% compared to another formulation. Even if this result was statistically significant, a mere improvement by just 1% is not considered clinically significant. After all, would you buy or use the mouth wash if it was only 1% better than a competitor's product? Probably not!

Physical Activity and Weight Gain Prevention

I-Min Lee, MBBS, ScD
Luc Djoussé, MD, DSc
Howard D. Sesso, ScD
Lu Wang, MD, PhD
Julie E. Buring, ScD

Context The amount of physical activity needed to prevent long-term weight gain is unclear. In 2008, federal guidelines recommended at least 150 minutes per week (7.5 metabolic equivalent [MET] hours per week) of moderate-intensity activity for "substantial health benefits."

Objective To examine the association of different amounts of physical activity with long-term weight changes among women consuming a usual diet.

Design, Setting, and Participants A prospective cohort study involving 34 079 healthy US women (mean age, 54.2 years) from 1992-2007. At baseline and months 36, 72, 96, 120, 144, and 156, women reported their physical activity and body weight. Women were classified as expending less than 7.5, 7.5 to less than 21, and 21 or more MET hours per week of activity at each time. Repeated-measures regression prospectively examined physical activity and weight change over intervals averaging 3 years.

Main Outcome Measure Change in weight.

Results Women gained a mean of 2.6 kg throughout the study. A multivariate analysis comparing women expending 21 or more MET hours per week with those expending from 7.5 to less than 21 MET hours per week showed that the latter group gained a mean (SD) 0.11 kg (0.04 kg; $P=.003$) over a mean interval of 3 years, and those expending less than 7.5 MET hours per week gained 0.12 kg (0.04; $P=.002$). There was a significant interaction with body mass index (BMI), such that there was an inverse dose-response relation between activity levels and weight gain among women with a BMI of less than 25 (P for trend $<.001$) but no relation among women with a BMI from 25 to 29.9 (P for trend $=.56$) or with a BMI of 30.0 or higher (P for trend $=.50$). A total of 4540 women (13.3%) with a BMI lower than 25 at study start successfully maintained their weight by gaining less than 2.3 kg throughout. Their mean activity level over the study was 21.5 MET hours per week (≈60 minutes a day of moderate-intensity activity).

Conclusions Among women consuming a usual diet, physical activity was associated with less weight gain only among women whose BMI was lower than 25. Women successful in maintaining normal weight and gaining fewer than 2.3 kg over 13 years averaged approximately 60 minutes a day of moderate-intensity activity throughout the study.

JAMA. 2010;303(12):1173-1179 www.jama.com

THE PREVALENCE OF OVERweight and obesity in the United States has increased dramatically over the past 2 decades, with 1 in 3 adults currently obese.[1] These numbers present a tremendous health care challenge in treatment and cost relating to the many adverse health conditions associated with excess body weight.[2,3]

At a fundamental level, weight gain occurs when caloric intake exceeds caloric expenditure. Many studies have examined physical activity, with or without caloric restriction, and weight loss among those who are overweight or obese.[4] Effective strategies exist for weight loss, but the majority of persons losing weight do not maintain their weight loss.[5,6] Because the average US adult gains weight with age,[7,8] developing ways to prevent unhealthful weight gain would help them avoid having to lose weight and then trying to maintain that loss. Compared with the

Physical Activity and Weight Gain Prevention

Results Women gained a mean of 2.6 kg throughout the study. A multivariate analysis comparing women expending 21 or more MET hours per week with those expending from 7.5 to less than 21 MET hours per week showed that the latter group gained a mean (SD) 0.11 kg (0.04 kg; $P=.003$) over a mean interval of 3 years, and those expending less than 7.5 MET hours per week gained 0.12 kg (0.04; $P=.002$). There was a significant interaction with body mass index (BMI), such that there was an inverse dose-response relation between activity levels and weight gain among women with a BMI of less than 25 (P for trend $<.001$) but no relation among women with a BMI from 25 to 29.9 (P for trend $=.56$) or with a BMI of 30.0 or higher (P for trend $=.50$). A total of 4540 women (13.3%) with a BMI lower than 25 at study start successfully maintained their weight by gaining less than 2.3 kg throughout. Their mean activity level over the study was 21.5 MET hours per week ($\approx$60 minutes a day of moderate-intensity activity).

JAMA, March 24/31, 2010—Vol 303, No. 12

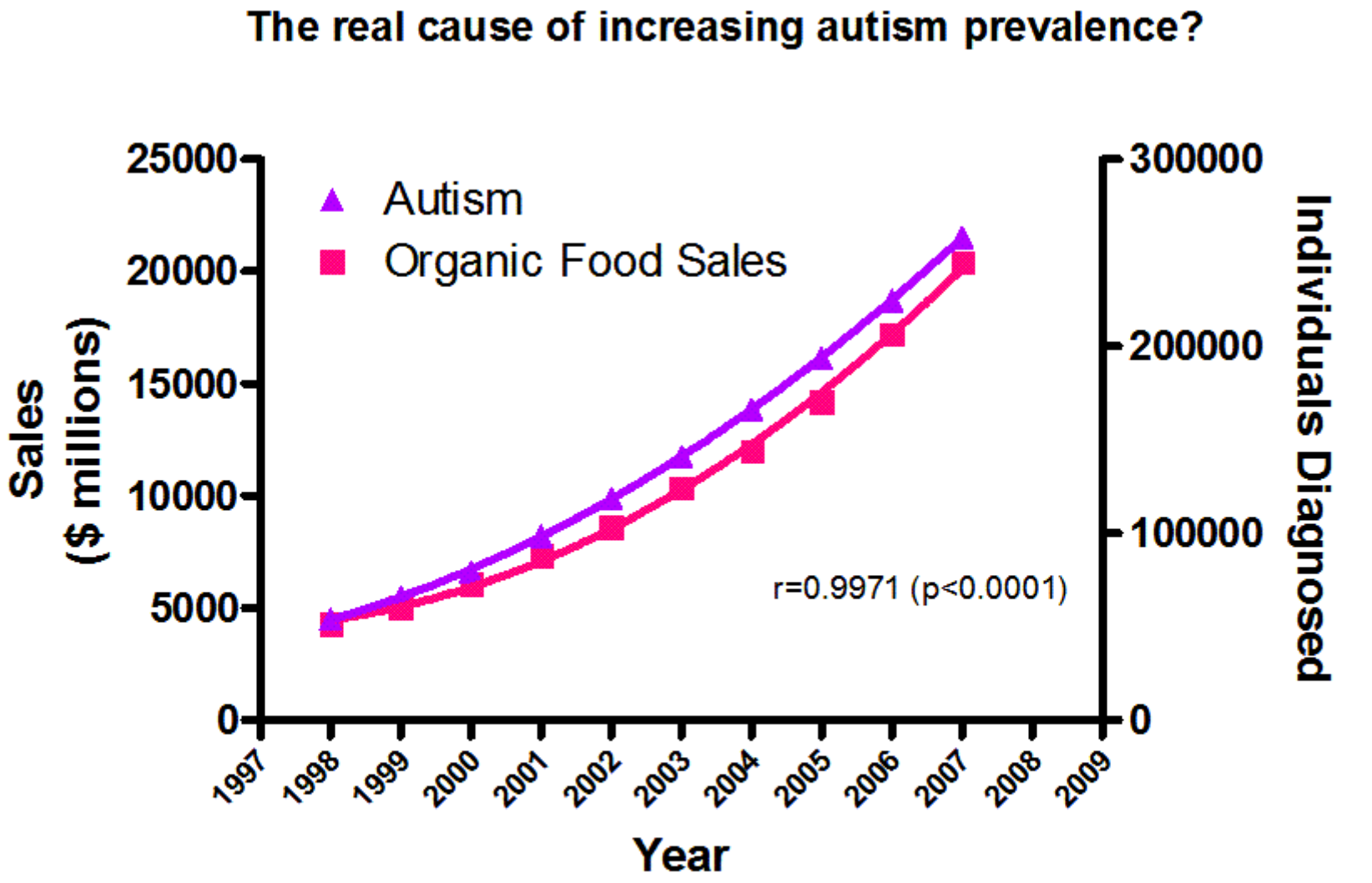
Physical Activity and Weight Gain Prevention

Table 2. Mean (SD) Differences in Weight Over Any 3-Year Period by Physical Activity Level, Women's Health Study, 1992-2007[a]

Group	No. of Women[b]	Physical Activity, MET Hours per Week			*P* Value for Trend	*P* Value for Interaction
		<7.5	7.5 to <21	≥21		
All women						
Analytical model[c]						
1		0.15 (0.04)	0.12 (0.04)	0 [Reference]	<.001	
2		0.12 (0.04)	0.11 (0.04)	0 [Reference]	<.001	
Age, y						
<55	21 363	0.12 (0.08)	0.02 (0.08)	0 [Reference]	<.001	<.001
55-64	9699	0.24 (0.06)	0.19 (0.06)	0 [Reference]	<.001	
≥65	3017	−0.09 (0.07)	0.07 (0.07)	0 [Reference]	.13	
BMI						
<25.0	17 475	0.21 (0.04)	0.14 (0.04)	0 [Reference]	<.001	<.001
25-29.9	10 516	−0.04 (0.06)	−0.04 (0.06)	0 [Reference]	.56	
≥30.0	6088	0.16 (0.14)	0.13 (0.16)	0 [Reference]	.50	
Smoking status						
Never	17 692	0.18 (0.05)	0.17 (0.05)	0 [Reference]	<.001	.53
Former	12 169	0.06 (0.06)	0.05 (0.06)	0 [Reference]	.04	
Current	4186	0.15 (0.15)	0.12 (0.16)	0 [Reference]	.11	
Menopausal status						
Premenopausal	9821	0.19 (0.13)	0.08 (0.13)	0 [Reference]	.03	.04
Postmenopausal	17 762	0.12 (0.04)	0.12 (0.04)	0 [Reference]	<.001	

JAMA, March 24/31, 2010—Vol 303, No. 12

Confusion of correlations, relationships, and causations



Sources: Organic Trade Association, 2011 Organic Industry Survey; U.S. Department of Education, Office of Special Education Programs, Data Analysis System (DANS), OMB# 1820-0043: "Children with Disabilities Receiving Special Education Under Part B of the Individuals with Disabilities Education Act

P-hacking

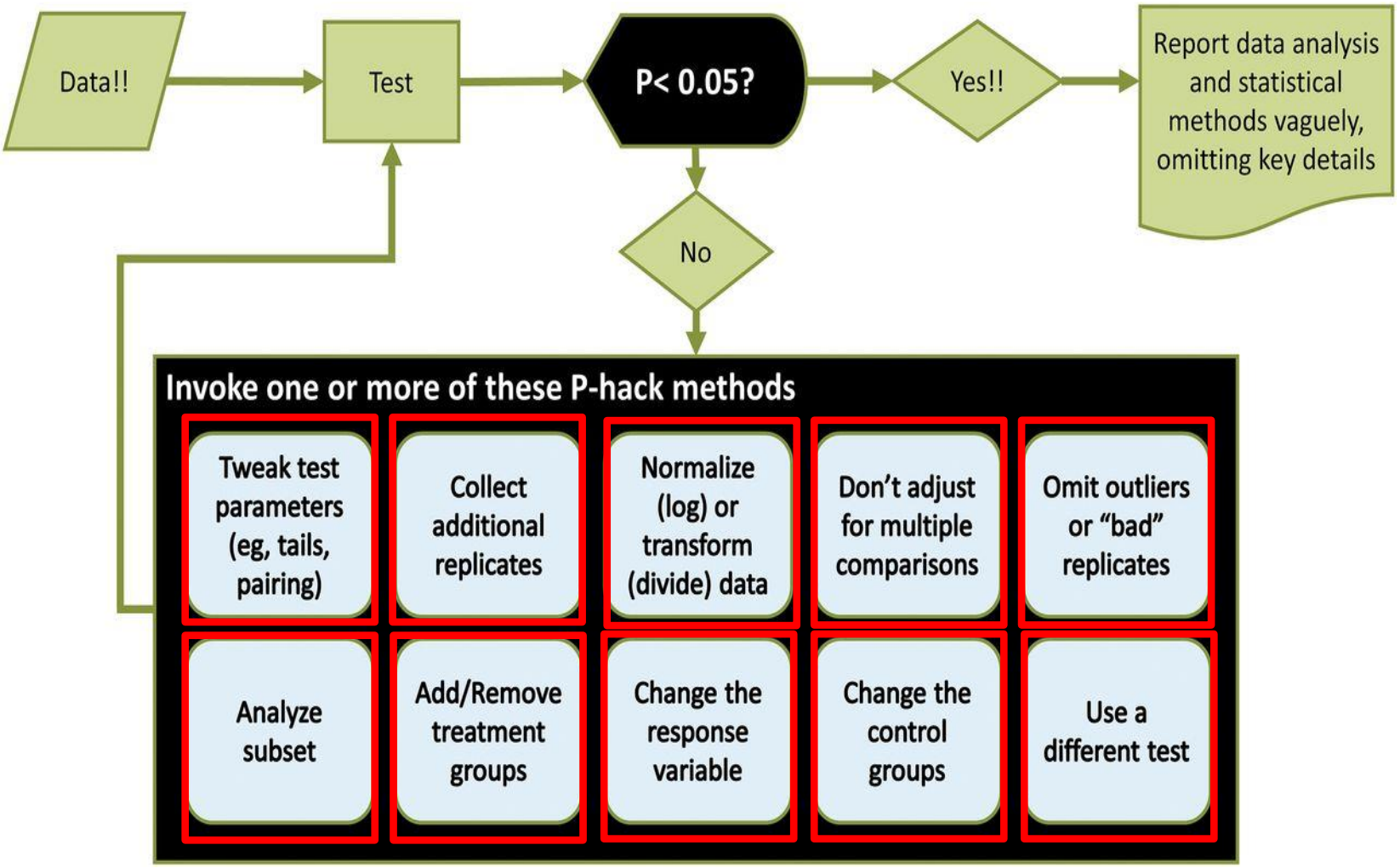


Fig 1. P-hacking refers to a series of analyses in which the goal is not to answer a specific scientific question but rather to find a hypothesis and data analysis method that results in a *P* value less than 0.05.

Molecular Pharmacology January 2020, 97 (1) 49-60

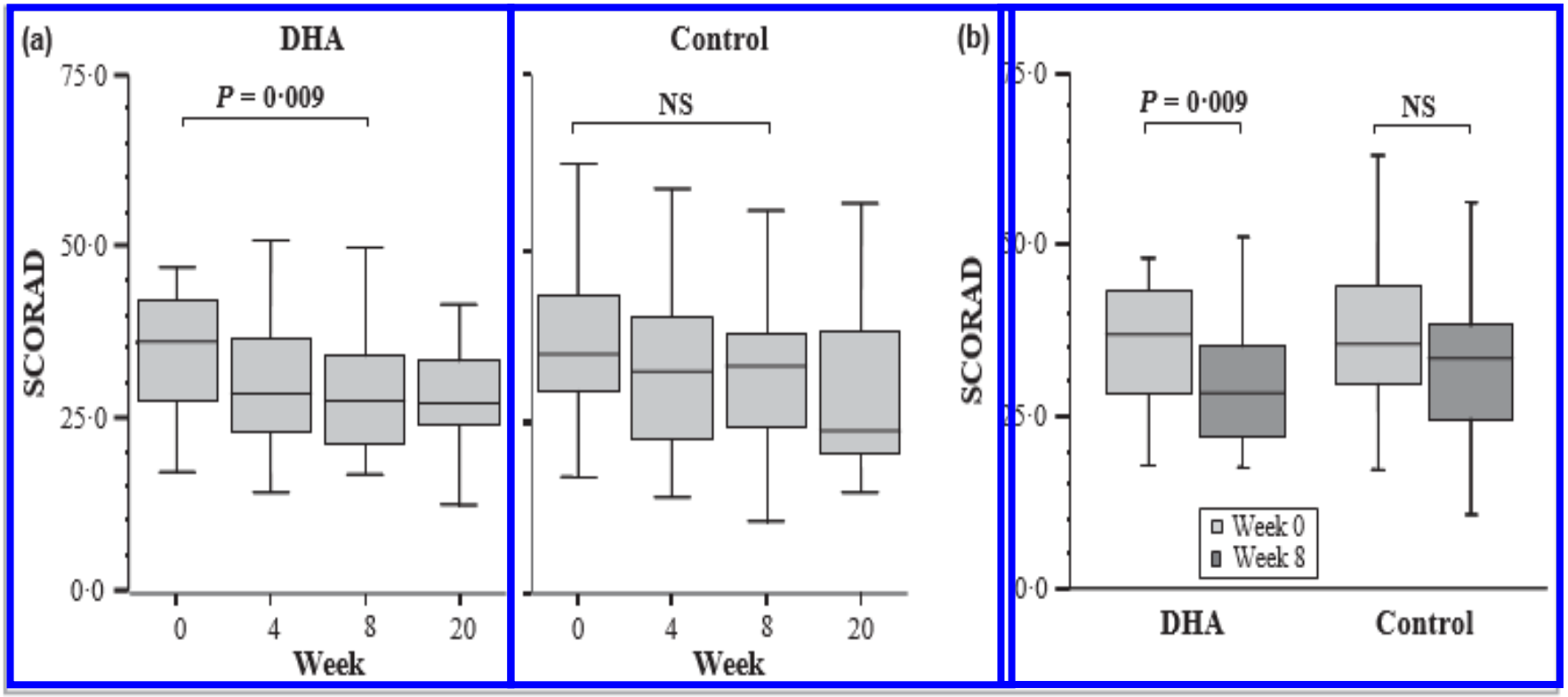
Statistics and the Relationship of Clinical Research to Clinical Practice

situations. As an example, consider a situation where an investigator finds significant results using the traditional approach, but when the *a priori* knowledge is examined, the posterior probability of effect may become much lower than the anticipated 95% using the Bayesian approach. In addition, the potential misuse of this approach is possible, as when findings do not achieve the 5% level of significance, tempting researchers to present their data in the Bayesian format. Moreover, substantial *a priori* knowledge may introduce potential ethical concerns in the conduct of trials when transitioning from Phase II to Phase III, whereas in studies using Bayesian approaches that is avoided by the independent replication of the frequentist approach.

Statistical Errors in Scientific Studies

1) Flawed and inadequate hypothesis
2) Improper study design
   A. Inadequate sample size
   B. Lack of adequate control condition/group
3) Overstatement of the analysis results
   A. Excessive interpretation of limited or insignificant results
   B. Confusion between P value and clinical significance
   C. p-hacking
   D. Confusion of correlations, relationships, and causations
4) Inappropriate presentation of the results and effects

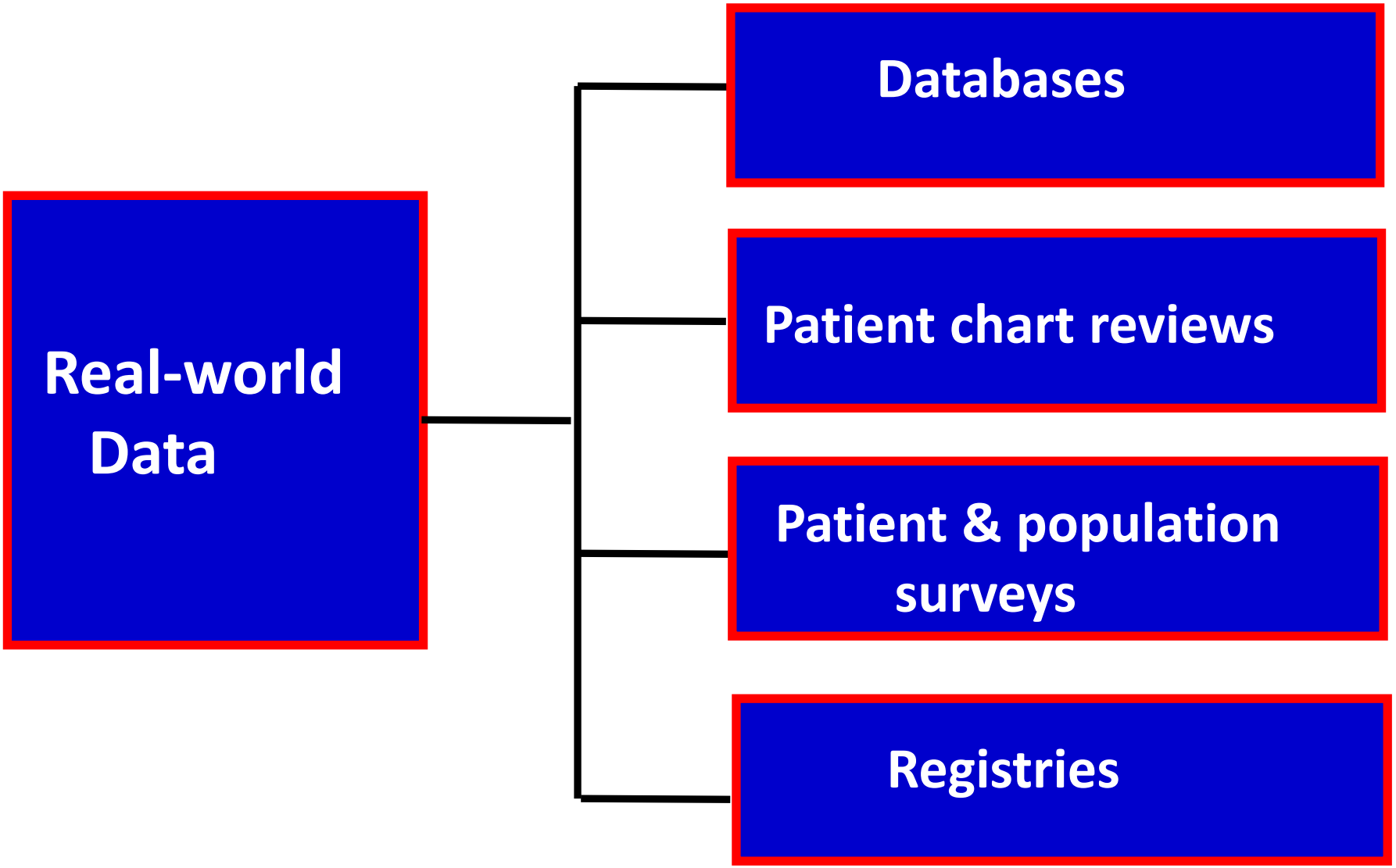
Misleading "significance comparisons"



- "the effect was significant in the treatment group, but not significant in the control group" does not imply that the groups differ significantly

British Journal of Dermatology 2008 158, pp786–792

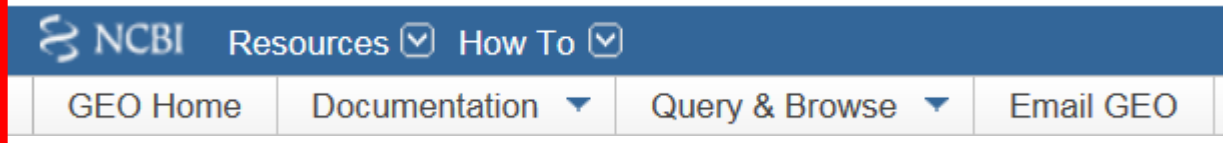
Sources of “real-world data”



Adapted from Ann Nutr Metab 2018;72(suppl 3):13–23

OMICS Data Bases

http://www.ncbi.nlm.nih.gov/geo/



Gene Expression Omnibus

https://www.ebi.ac.uk/arrayexpress/

Series GSE93606

Query DataSets for GSE93606

Status	Public on Jan 14, 2017
Title	Host-Microbial interactions in Idiopathic Pulmonary Fibrosis
Organism	Homo sapiens
Experiment type	Expression profiling by array

Gene Expression Omnibus

Summary

Changes in the respiratory microbiome are associated with disease progression in Idiopathic pulmonary fibrosis (IPF). The role of the host response to the respiratory microbiome however remains unknown. The role of this study is to explore the host-microbial interaction in IPF. Network analysis of gene expression data identified two gene modules that strongly associate with a diagnosis of IPF, BAL bacterial burden (determined by 16S quantitative PCR) and specific microbial OTUs, as well as lavage and peripheral blood neutrophilia. Genes within these modules that are involved in the host defence response include NLRC4, PGLYRP1, MMP9, DEFA4. The modules also contain two genes encoding specific antimicrobial peptides (SLPI and CAMP). Many of these particular transcripts were associated with survival and showed longitudinal over expression in subjects experiencing disease progression, further strengthening their relationship with disease. Integrated analysis of the host transcriptome and microbial signatures demonstrates an apparent host response to the presence of an altered or more abundant microbiome. These responses remain elevated on longitudinal follow up, suggesting that the bacterial communities of the lower airways may be acting as persistent stimuli for repetitive alveolar injury in IPF.

Sixty patients diagnosed with IPF were prospectively enrolled, together with 20 matched controls. Subjects underwent bronchoalveolar lavage (BAL) and peripheral whole blood was collected into PAXgene tubes for all subjects at baseline. For IPF subjects additional samples were taken at 1, 3, and 6 months and (if alive) a year. Gene expression profiles were generated using Affymetrix Human Gene1.1ST Arrays.

Overall design

Survival=Months from Recruitment to composite end point or censoring; Age=Age in years at recruitment; FVC= Percent predicted Forced Vital Capacity; DLCO=Percent predicted Diffusing capacity of the lungs for carbon monoxide; Composite_End_Point=Death or decline in FVC >10% over a six month period, 1=Event, 2=No event.

Sample GSM2458605

Status	Public on Jan 14, 2017
Title	IPF_1008, Timepoint 0
Sample type	RNA
Source name	whole blood
Organism	Homo sapiens
Characteristics	tissue: whole blood disease state: Idiopathic Pulmonary Fibrosis gender: male survival (months): 11 age (years): 69 fvc: 66.7 dlco: 29 composite_end_point: 1

NCBI

Gene Expression Omnibus

Series GSE93606

Host Microbial interaction in Idiopathic pulmonary Fibrosis

Accession	Disease state	Gender	Characteristics	Fvc	Dlco	CEP
GSM2458563	Control	male	survival (months): 34 age (years): 71			0
GSM2458569	Control	male	survival (months): 34 age (years): 51			0
GSM2458580	Control	male	survival (months): 34 age (years): 70			0
GSM2458582	Control	male	survival (months): 34 age (years): 66			0
GSM2458583	IPF	female	survival (months): 10 age (years): 57	99.2	23.4	1
GSM2458586	IPF	male	survival (months): 34 age (years): 65	67.1	37.9	0
GSM2458591	IPF	male	survival (months): 12 age (years): 75	58.7	39.1	1
GSM2458596	IPF	male	survival (months): 3 age (years): 58	43.3	29.6	1
GSM2458645	IPF	female	survival (months): 19 age (years): 65	70.9	27	1

Host Microbial interaction in Idiopathic pulmonary Fibrosis

Accession	X7892501	X7892502	X7892503	X7892504	X7892505
GSM2458563	1.2678318	3.25496289	1.6405874	7.1984219	2.226013
GSM2458564	1.604042	1.91023974	2.8257369	7.4804659	1.564427
GSM2458565	1.8529631	2.78357227	2.16909778	8.2231366	1.879694
GSM2458566	1.2019307	2.84519549	2.37242555	7.9281866	1.955655
GSM2458579	2.0986949	2.52552128	2.71025303	8.2800804	2.025326
GSM2458580	1.2699685	2.43952097	2.67639744	7.8051862	1.339459
GSM2458581	1.2232135	3.21436441	3.10102964	8.3806826	1.751142
GSM2458582	2.4210534	2.41303239	2.84733611	7.5971641	1.837432
GSM2458583	2.9944441	2.63793839	1.5628737	7.7759946	1.933745
GSM2458584	2.3035022	2.35793595	2.33795155	7.8556433	2.572528
GSM2458585	2.1916808	3.2682914	1.85568322	7.8108678	2.501918
GSM2458586	1.5482853	4.03812911	2.72804728	7.3792546	1.440081
GSM2458587	2.2205105	2.23889942	2.42653737	7.2436966	1.484616

Table 2. The Top 20 Transcript Clusters Significant at a 1% False Discovery Rate Ordered by Fold Change

Gene Name	Avg Expr	*t* Statistic	*P* Value	B-H–Adjusted *P* Value	Absolute Fold Change
ORM1	5.56	6.68	2.79×10^{-9}	3.61×10^{-6}	3.62
DEFA4	6.62	3.69	0.0004	0.0051	3.04
CD177	5.66	4.63	1.39×10^{-5}	0.0006	2.52
ARG1	5.09	4.02	0.0001	0.0027	2.29
SLPI	7.69	5.56	3.41×10^{-7}	7.05×10^{-5}	2.29
MMP9	7.75	4.79	7.42×10^{-6}	0.0004	2.28
RNASE3	6.49	4.17	7.61×10^{-5}	0.0019	2.26
TXN	7.80	8.80	1.96×10^{-13}	2.75×10^{-9}	2.22
BCL2A1	6.50	6.03	4.58×10^{-8}	2.01×10^{-5}	2.19
TNFAIP6	6.85	4.93	4.37×10^{-6}	0.0003	2.15
SNORD64	4.68	−4.86	5.57×10^{-6}	0.0003	−2.11
ANXA3	7.23	5.38	7.12×10^{-7}	0.0001	2.11
CAMP	7.24	4.78	7.82×10^{-6}	0.0004	2.11
CSTA	8.41	8.48	8.27×10^{-13}	5.81×10^{-9}	2.06
HP	5.51	4.11	9.29×10^{-5}	0.0021	2.05
CLEC4D	5.87	5.18	1.58×10^{-6}	0.0001	2.02
SUB1	7.32	5.37	7.23×10^{-7}	0.0001	2.01
OLFM4	4.80	2.88	0.005	0.0301	2.00
PGLYRP1	7.64	5.72	1.71×10^{-7}	4.54×10^{-5}	1.99
RPL26	8.92	5.64	2.42×10^{-7}	5.66×10^{-5}	1.97

Definition of abbreviations: Avg Expr = average $\log_2$-adjusted expression level for that gene across all the arrays; B-H = Benjamini-Hochberg. The highest fold change in complete set of differentially expressed genes (n = 1,358) was 3.62.

Electronic Health Records

ID	Hospital	Admit	Discharge	Age	Gender	Ethnicity	BMI	HTN	HPL	CKD	COPD	Asthma	OSA	Cirrhosis	PUD	CAD	CHF
307	3	04/02/2020	04/11/2020	47	1	2	31.87	0	0	0	1	0	0	0	0	0	0
310	3	04/02/2020	04/12/2020	48	1	1	47.47	1	1	0	0	0	1	0	0	1	0
313	3	04/03/2020	04/09/2020	69	0	1	27.79	1	1	0	0	0	0	0	0	0	0
317	3	04/03/2020	04/16/2020	63	1	2	30.51	1	1	0	0	0	0	0	0	0	0
411	4	43935	43938	65	0	2	35	1	0	0	0	1	0	0	0	0	0
416	4	43919	43921	49	1	2	33	1	0	0	0	0	0	0	0	0	0
455	4	43950	43952	20	0	1	34	0	0	0	0	1	0	0	0	0	0
476	4	43959	43972	54	1		37	1	1	1	0	0	1	0	0	1	0
522	5	43918	43924	44	0		34.1	0	0	0	0	0	0	0	0	0	0
532	5	43921	43928	61	0	2	47.6	1	1	0	0	0	0	0	0	0	0
543	5	43925	43928	27	1	4	36.8	0	0	0	0	1	0	0	0	0	0
572	5	43938	43950	67	0	1	47.6	1	1	1	1	0	1	0	1	1	1
573	5	43939	43943	61	1	2	28.2	1	1	0	0	0	0	0	0	0	0
575	5	43940	43948	57	0	2	27.4	0	0	0	0	0	0	0	0	0	0
580	5	43940	43945	53	1	1	50.2	1	1	0	0	0	0	0	0	0	0
611	5	43955	43967	37	1	2	37.3	0	0	0	0	0	0	0	0	0	0
622	5	43960	43972	84	1	1	24.4	1	1	1	0	0	0	0	0	1	1
632	5	43978	6/1/20200	68	1	1	26.1	0	0	0	0	0	0	0	0	0	0

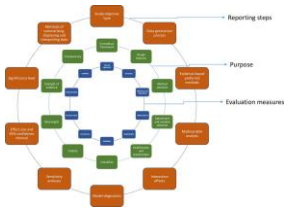
Quality Assurance vs. Research

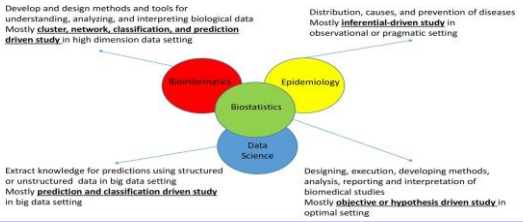
- In general, a quality assurance project is a project that is focused primarily on improving patient care within a given patient care environment and, as such, the outcome may not be generalizable to other patient care environments.
- There is usually a commitment, in advance of data collection, to a corrective plan given any one of a number of study outcomes.
- The study lacks:
  1) Prospective assignment of patients to different procedures or therapies based on a predetermined plan
  2) Control group" in whom the therapeutic or study intervention is intentionally withheld to allow an assessment of its efficacy?

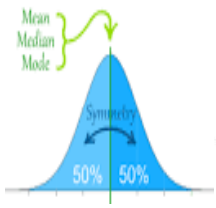
Statistical Support

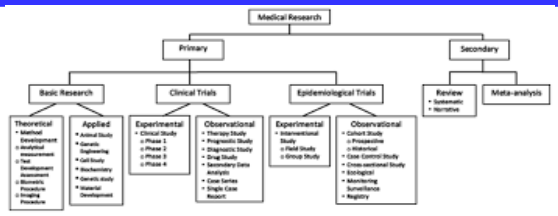
- Study Design
- Statistical Analysis
- Consultation on
  - Report writing
  - “revise and resubmit”
  - Other

THANK YOU


Reporting steps
Purpose
Evaluation measures


Develop and design methods and tools for understanding, analyzing, and interpreting biological data
Mostly cluster, network, classification, and prediction driven study in high dimension data setting
Distribution, causes, and prevention of diseases
Mostly inferential-driven study in observational or pragmatic setting
Bioinformatics
Epidemiology
Biostatistics
Data Science
Extract knowledge for predictions using structured or unstructured data in big data setting
Mostly prediction and classification driven study in big data setting
Designing, execution, developing methods, analysis, reporting and interpretation of biomedical studies
Mostly objective or hypothesis driven study in optimal setting


Mean
Median
Mode
Symmetry
50%
50%


Medical Research
Primary
Secondary
Basic Research
Clinical Trials
Epidemiological Trials
Review
Meta-analysis
Theoretical
Applied
Experimental
Observational
Experimental
Observational