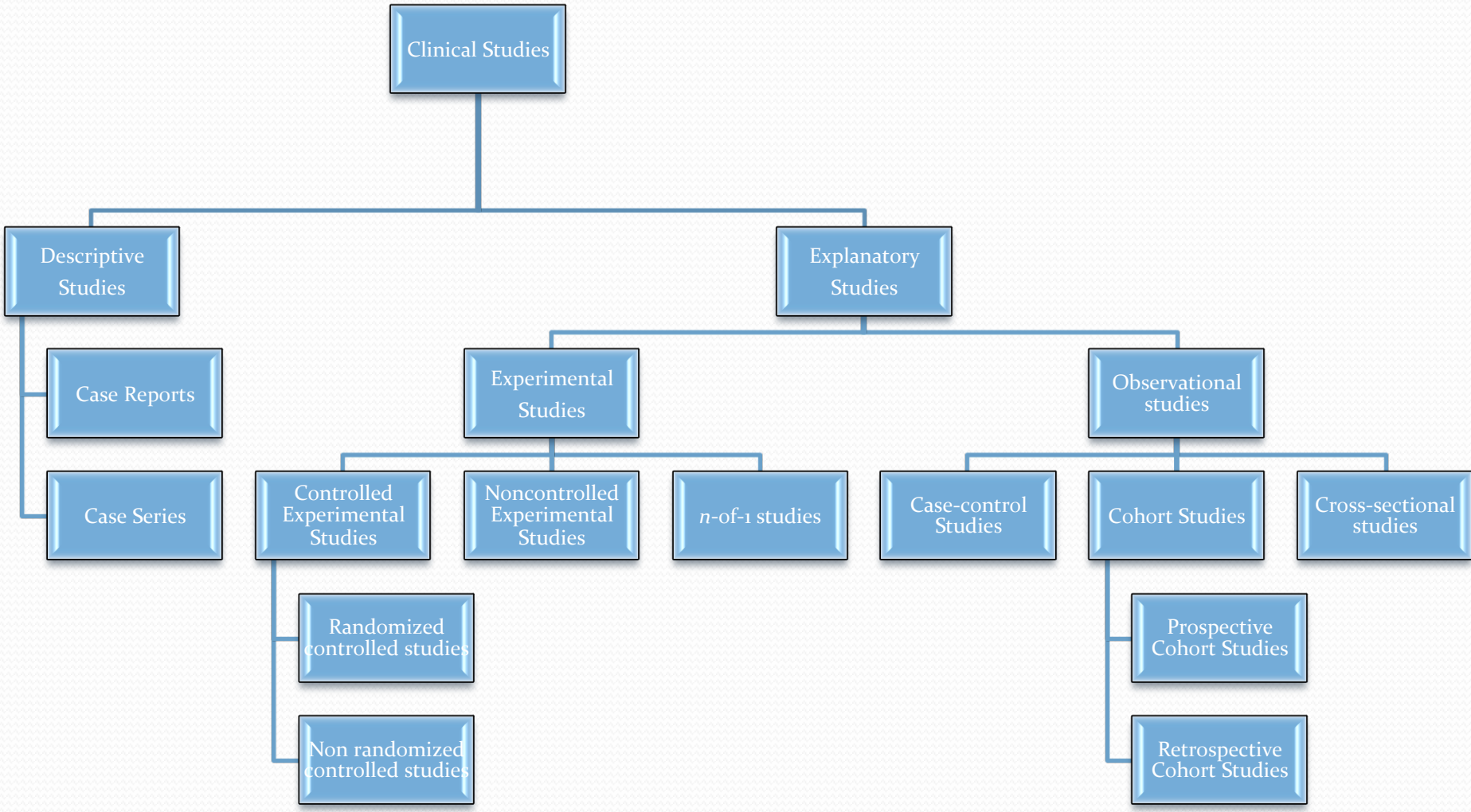
# Evaluating clinical studies

Steps in evaluating clinical studies

- Step 1: What type of study is it?
- Step 2: The journal, authors, and study purpose
- Step 3: Methods used
- Step 4: Statistical analysis
- Step 5: Results, interpretation and conclusion
- Step 6: Putting it all together

- There are 2 types of studies
  - Descriptive
    - Simply recording information from observing patients
  - Explanatory
    - Using group comparisons as the basis for determining whether an exposure/treatment might cause or affect a condition or outcome


Clinical Studies
Descriptive Studies
Explanatory Studies
Case Reports
Case Series
Experimental Studies
Observational studies
Controlled Experimental Studies
Noncontrolled Experimental Studies
n-of-1 studies
Case-control Studies
Cohort Studies
Cross-sectional studies
Randomized controlled studies
Non randomized controlled studies
Prospective Cohort Studies
Retrospective Cohort Studies

# Descriptive studies

- Descriptive studies are not generally considered to be studies and are referred to as reports
  - Case reports: Reporting observations in one or a small number of individual patients
  - Case series Reporting observations from a small group or series of patients

# Explanatory studies

- Explanatory studies
  - Experimental studies
    - Controlled
    - Noncontrolled
  - Observational studies

# Experimental studies

- Involve actual intervention by investigators
- Subjects are assigned and given treatments by investigators
- Controlled experimental studies are the best “gold standard”
  - They use a treatment group and a control group

# Control group

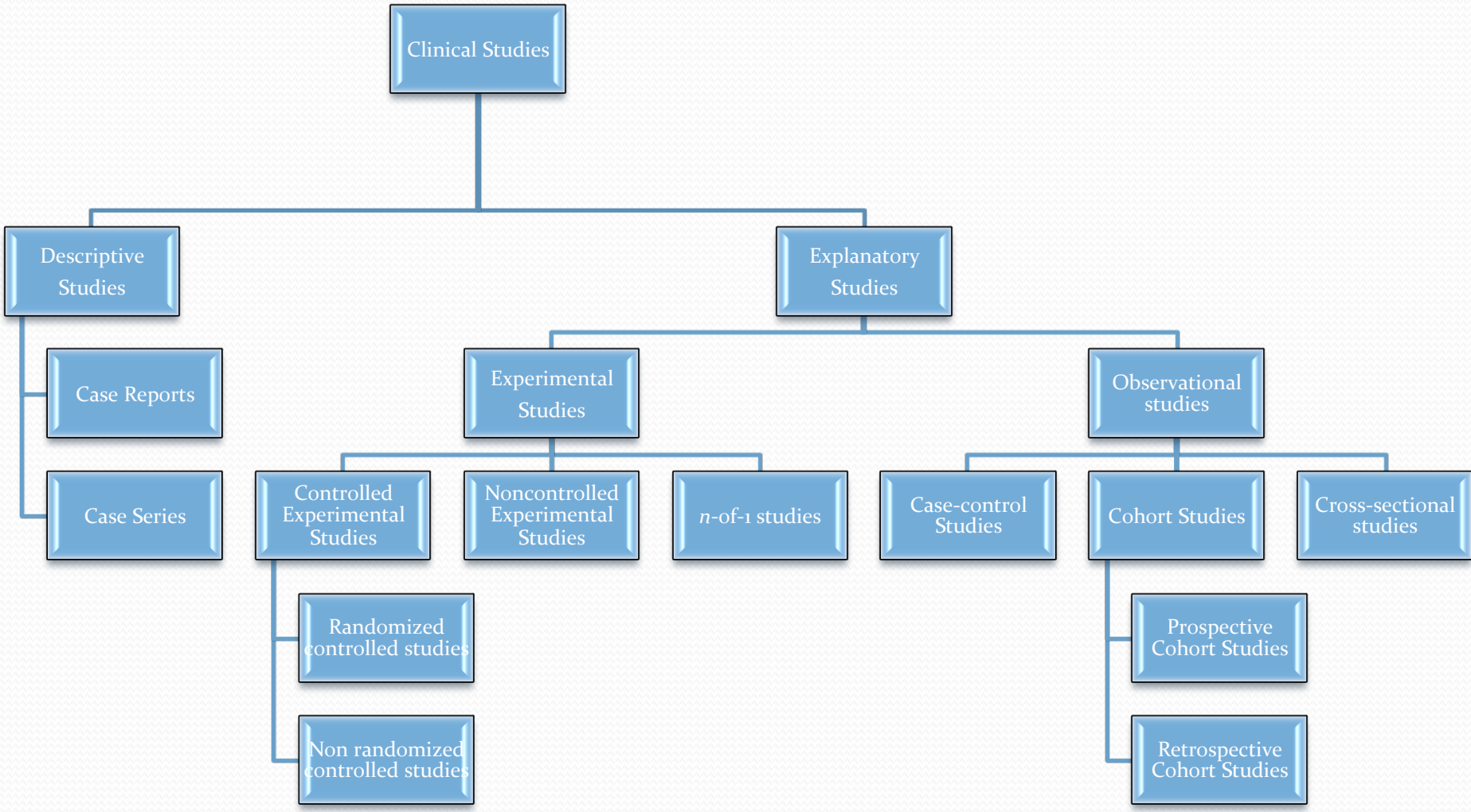
- Helps account for factors other than treatment that might affect the study results
- Investigators compare effects seen in the control patients with those in the treatment patients to determine if there is a difference between them
- Types of control groups
  - Placebo
  - Active (another treatment with established efficacy)
  - No treatment
  - Historical (Comparison with a treatment previously studied)
    - Not commonly used, only when it is the only type of control available

# *n*-of-1 studies

- Type of experimental studies
- Single-subject research design
- Often used by primary care practitioners
- Studies a specific patient
  - The researcher conducts a baseline assessment of the patient's condition followed by therapy initiation
  - During/after therapy the researcher measures changes in the condition

# *n*-of-1 studies

- Disadvantages of n-of-1 studies
  - Inability to generalize results to others
  - Difficult/impossible to perform statistical analyses
  - Difficult to validate studies


Clinical Studies
Descriptive Studies
Explanatory Studies
Case Reports
Case Series
Experimental Studies
Observational studies
Controlled Experimental Studies
Noncontrolled Experimental Studies
n-of-1 studies
Case-control Studies
Cohort Studies
Cross-sectional studies
Randomized controlled studies
Non randomized controlled studies
Prospective Cohort Studies
Retrospective Cohort Studies

# Observational studies

- In observational studies the treatment(s) taken or other exposures studied were *not given by* the study investigators

- Observational designs are used when controlled experimental study design is not possible, feasible or ethical
  - e.g., for rare conditions or those that require a long time to develop

# Example

**Coffee consumption and pancreatic cancer**

- Is coffee intake associated with an increased risk of pancreatic cancer development?
- Investigators suspect that coffee might be a risk factor for pancreatic cancer.
- **Would it be appropriate for the investigators to use an experimental design to test their hypothesis?**

# Case-control studies

- Used to determine the possible factors (e.g., exposures, drugs) influencing or causing an event or outcome.
- Always retrospective
- This design begins with patients who already have the event or outcome (cases) and another group of similar patients who lack the event or outcome (controls)
- The investigators need to look back in time in order to compare drug use or the extent of exposure in both groups prior to when they developed the outcome
- If the cases are found to have significantly greater drug use or extent of exposure than the controls, a possible association exists between the drug/exposure and outcome development

| | | |
|---|---|---|
| Case-control | **At start**<br><br>Select:<br><br>Cases: patients who already have condition or outcome being studied<br><br>Controls: patients who do not have the condition or outcome being studied but are otherwise similar to cases<br><br>**Compare in the groups**:<br><br>Extent of exposure to drugs or other factors thought to affect development of the condition or outcome: done *retrospectively* (looking back in past) through use of surveys, interviews, medical records, or medical databases<br><br>**Analyze and determine**:<br><br>Were cases more likely than controls to have been exposed to drugs or other possible causes of the condition/outcome? | A: Good for studying possible causes of adverse events or negative outcomes, especially rare/infrequent outcomes or those that take a long time to develop; faster, less expensive than prospective study designs<br><br>D: Possible selection bias (are cases truly comparable to controls?); recall bias (patients' memories of an exposure might be inaccurate); interviewer/observer bias (interviewer might slant data collection if aware of who is a case or control); records or databases could be inaccurate or incomplete |

# Cohort studies

- A cohort study can be prospective (concurrent) or retrospective (nonconcurrent, historical)
- The basic design of each is the same:

  (1) Identify groups (cohorts) with and without the drug use/ exposures of interest – no one has the outcome at the start

  (2) Follow the groups forward over time and measure differences in outcome development
- The nonconcurrent or retrospective design differs from the prospective cohort study in that all information (drug use/ exposures and outcomes) is obtained from already existing medical records or databases
- The start of a nonconcurrent cohort study occurs at a designated point in the past

# Cohort studies

- The investigators initially select the cohorts for inclusion in either the study or control groups with no knowledge of whether or not the outcome later develops
- Once all subjects are included the investigators examine the existing data, going forward in time from the starting point, to determine whether or not the subjects in each group developed the outcome

**Which cohort design, prospective or retrospective, is strongest?**

- The prospective concurrent design is best because it is less subject to bias and inaccuracies
- The nonconcurrent or retrospective design is dependent upon existing records or databases that might be incomplete or incorrect

# Cohort design

- Follows a study "cohort" (a group of individuals/ subjects who share a common characteristic) over time to determine if a drug or other exposure will lead to the development of an outcome of interest
- Unlike the case-control design, the subjects in a cohort study do not have the outcome at the start of the study.
- Investigators identify subjects who are taking the drug or have the exposure of interest (study subjects), as well as similar subjects who are not taking the drug or who lack the exposure (control/comparison subjects). The investigators then follow the subjects
- In both groups (through scheduled visits, by examining medical records) over a certain period of time to compare the extent to which they develop the outcome
- If significantly more subjects in the study group develop the outcome compared to the control subjects, it is concluded that the drug or exposure might contribute to outcome development.

| | | |
|---|---|---|
| Cohort (follow-up) | **At start**<br><br>Select:<br><br>Study subjects: subjects already taking certain drug(s) or who have an exposure(s) that might affect the outcome of interest<br><br>Control/comparison subjects: subjects who are not taking the drug(s) or who do not have the exposure, but who are otherwise similar to study subjects<br><br>**Compare in the groups**:<br><br>Extent to which subjects in each group develop the condition/outcome: done by following subjects forward over time (*prospectively*, if starting point in present time and data about outcome development obtained in future; *retrospectively*, if complete medical records exist over an extended time that allow the starting point (exposure or no exposure) to be in the past, with data about outcome development obtained by looking through subsequent medical records) | A: Good for studying possible causes of adverse events or outcomes especially if they occur relatively commonly, or how/if certain exposures or characteristics might affect later outcome development<br><br>D: Possible selection bias (are study subjects truly comparable to control/comparison subjects?); with prospective design: can take long time to complete, subject drop-out (loss to follow-up) could occur, potential expense; retrospective design shares disadvantages of case-control design |

# Cross-sectional design

- The study sample is selected from a targeted population of interest and information about both the extent of drug use/other exposures and presence of the outcome is obtained from the sample at the same time
- Provides a "cross-section" snapshot of the prevalence or existence of specific conditions, characteristics, and outcomes at one point in time
- The investigators obtain all the exposure and outcome information from the study sample through the use of questionnaires or surveys
- The data from subjects within the sample are compared and analyzed based on the presence or absence of these factors
- Limitations are similar to case-control study as they collect data about past exposures or drug use from subjects' recollections or records
- The cross-sectional study lacks a separate control/comparison group

| | | |
|---|---|---|
| Cross-sectional (prevalence, survey) | **At start**:<br><br>Subjects identified who represent a population of interest; they might/might not have the exposure(s) or the outcome(s) of interest<br><br>**Compare in group**:<br><br>Presence and absence of both exposures and the condition or outcome of interest: determined at same time, resulting in "snapshot" or "cross-section" of the exposures and outcomes present at one given time (done through use of surveys, questionnaires, or examination of databases or patient records)<br><br>**Analyze and determine**:<br><br>What are the factors or exposures associated with certain conditions or outcomes, and what is their prevalence? | A: Can identify possible risk factors for or potential causes of a disease or condition, the prevalence of a disease or condition at a specific point in time, or if beliefs or practices affect health or behavioral outcomes; relatively quick, inexpensive<br><br>D: Cannot determine whether the "cause" (characteristics or exposures) actually preceded or resulted in the "effect" (outcomes); group identified for study inclusion might not adequately represent desired population of interest; subject self-reporting in surveys or questionnaires might not be accurate or nonresponse might be a problem |

# Which study is strongest?

- The order of the study designs from strongest (best) to weakest (most limitations/disadvantages) is:

1. Controlled experimental
2. Prospective cohort
3. Case-control/cross-sectional/retrospective cohort

- Observational studies cannot prove that a drug or exposure caused a certain outcome; only well-designed controlled experimental studies can do this
- Observational studies can provide very useful information when it is not possible, feasible, or ethical to conduct an experimental study

# Bias in clinical trials

- "Bias refers to unconscious distortion in the selection of patients, collection of data, determination of end points, and final analyses" (Shapiro & Louis, 1983)

- Bias is also referred to as systematic error, and can be defined as: "Any process or effect at any stage of a study from its design to its execution to the application of information from the study, that produces results or conclusions that differ systematically from the truth" (**Gay, 1999**)

# Types of Bias

- Selection bias: it is related to the recruitment of subjects into different groups with unusual and unequal relation
- Drop-out bias (loss to follow-up): occur when a subject leaves a trial before it's over.
- Information and Misclassification bias: result from error in measuring outcome or exposure that results in differential accuracy of information between compared groups
- Confounding: occurs when a risk factor affecting health status or outcome is not considered

  Example: confounding by reason for prescription; and confounding by co-medication

# Types of Bias

- Bias due to lack of compliance
- Publication bias: it is caused by the tendency of publishing studies with positive results rather than negative
- Bias due to tendency toward obtaining positive results, frequently patients like getting and reporting positive results, similarly, investigators and statisticians wish to see the drug approved, especially if they have financial interest with the company developing the drug under investigation

# Evaluating clinical studies

- Journal
- Authors
- Study purpose

- There are thousands of journals that vary in quality
- Editorial boards and peer review are 2 methods for ensuring the overall quality of a journal and its studies

# Journals

- Editorial board
  - Consists of individuals with expertise in the journal's area of focus
  - Helps assure the quality of the published studies
  - The editors read the manuscript first and make the decision to send it for peer review or not
- Peer review
  - The manuscript is sent by the editor to a small number of outside individuals (peers) with expertise in the subject area
  - The peers provide their comments/ revisions/ recommendations about accepting or rejecting the manuscript

- How to determine if a published study was peer reviewed or not?
  - Check the journal instructions for authors
  - Check the received date and date of acceptance

- Investigators should conduct their study in a manner free from bias or other factors that can affect their objective judgment
- Competing or conflicting interests can influence the manner by which investigators conduct the study or view the results
- Conflict of interest for investigators can compromise the objectivity and quality of their work
- Conflict of interests can be personal or financial (ties to companies, funding, etc., easier to identify)
- Any conflict of interest should be clearly stated in the publication
- Conflicts of interest do not necessarily invalidate the study, they indicate that readers should use extra care when analyzing the study

- Potential conflicts of interest
  - Receiving study funding from the manufacturer of the drug investigated
  - Serving as a consultant or on the board of directors of the pharmaceutical manufacturer of the drug investigated
  - Being employed by the manufacturer of the drug studied
  - Having a personal relationship or representing the manufacturer of the drug under investigation
- If a pharmaceutical manufacturer only provided the drug or placebo used in the study without any other involvement, it is not a conflict of interest

- Questions to determine if conflicts of interests exist
  - Did the introduction appear overly positive or only focus on the benefits of therapy?
  - Inclusion or exclusion criteria that include patients who are more likely to benefit from treatment?
  - Was the active control chosen so that clinicians would choose the investigated drug?
  - Were there any conclusions that are not supported by the results in the study?

- Things to consider when critically reading an article
  - Journal quality (Impact factor…)
  - Potential conflict of interests
  - The objectives of the study and the related hypotheses to determine if the design and methods were sufficient to fulfill the purpose

InCites™ Journal Citation Reports®

THOMSON REUTERS

**Journal Data Filtered By: Selected JCR Year: 2014 Selected Editions: SCIE,SSCI**

| Rank | Full Journal Title | Total Cites | Journal Impact Factor |
|---|---|---|---|
| 1 | CA-A CANCER JOURNAL FOR CLINICIANS | 18,594 | 115.840 |
| 2 | NEW ENGLAND JOURNAL OF MEDICINE | 268,652 | 55.873 |
| 3 | CHEMICAL REVIEWS | 137,600 | 46.568 |
| 4 | LANCET | 185,361 | 45.217 |
| 5 | NATURE REVIEWS DRUG DISCOVERY | 23,811 | 41.908 |
| 6 | NATURE BIOTECHNOLOGY | 45,986 | 41.514 |
| 7 | NATURE | 617,363 | 41.456 |
| 8 | Annual Review of Immunology | 16,750 | 39.327 |
| 9 | NATURE REVIEWS MOLECULAR CELL BIOLOGY | 35,928 | 37.806 |
| 10 | NATURE REVIEWS CANCER | 39,868 | 37.400 |
| 11 | NATURE REVIEWS GENETICS | 29,388 | 36.978 |
| 12 | NATURE MATERIALS | 64,622 | 36.503 |
| 13 | JAMA-JOURNAL OF THE AMERICAN MEDICAL | 126,479 | 35.289 |
| 14 | NATURE REVIEWS IMMUNOLOGY | 28,938 | 34.985 |
| 15 | Nature Nanotechnology | 34,387 | 34.048 |

| | | | |
|---|---|---|---|
| 16 | SCIENCE | 557,558 | 33.611 |
| 17 | CHEMICAL SOCIETY REVIEWS | 81,907 | 33.383 |
| 18 | Annual Review of Astronomy and Astrophysics | 8,462 | 33.346 |
| 19 | Nature Photonics | 23,499 | 32.386 |
| 20 | CELL | 201,108 | 32.242 |
| 21 | NATURE METHODS | 32,342 | 32.072 |
| 22 | NATURE REVIEWS NEUROSCIENCE | 32,989 | 31.427 |
| 23 | Annual Review of Biochemistry | 19,927 | 30.283 |
| 24 | REVIEWS OF MODERN PHYSICS | 39,402 | 29.604 |
| 25 | NATURE GENETICS | 85,481 | 29.352 |
| 26 | PROGRESS IN MATERIALS SCIENCE | 8,475 | 27.417 |
| 27 | NATURE MEDICINE | 62,572 | 27.363 |
| 28 | PHYSIOLOGICAL REVIEWS | 24,528 | 27.324 |
| 29 | PROGRESS IN POLYMER SCIENCE | 19,454 | 26.932 |
| 30 | Nature Chemistry | 16,973 | 25.325 |
| 31 | LANCET ONCOLOGY | 24,861 | 24.690 |
| 32 | NATURE REVIEWS MICROBIOLOGY | 18,866 | 23.574 |
| 33 | CANCER CELL | 27,283 | 23.523 |
| 34 | Annual Review of Plant Biology | 16,494 | 23.300 |
| 35 | LANCET INFECTIOUS DISEASES | 13,161 | 22.433 |
| 36 | ACCOUNTS OF CHEMICAL RESEARCH | 53,349 | 22.323 |
| 37 | Cell Stem Cell | 17,720 | 22.268 |
| 38 | TRENDS IN COGNITIVE SCIENCES | 20,396 | 21.965 |

| | | | |
|---|---|---|---|
| 11719 | Journal of Meteorological Research | 5 | Not Available |
| 11719 | Journal of Oral & Facial Pain and Headache | 14 | Not Available |
| 11719 | Journal of Transport & Health | 49 | Not Available |
| 11719 | Journal of Vibration Engineering & Technologies | 5 | Not Available |
| 11719 | Journal of the American Association of Nurse Practitioners | 48 | Not Available |
| 11719 | Journal of the American Association of Nurse Practitioners | 48 | Not Available |
| 11719 | Journal of the Association for Information Science and Technology | 78 | Not Available |
| 11719 | Journal of the Association for Information Science and Technology | 78 | Not Available |
| 11719 | Lancet Psychiatry | 81 | Not Available |
| 11719 | Lancet Psychiatry | 81 | Not Available |
| 11719 | Language Cognition and Neuroscience | 38 | Not Available |
| 11719 | Language Cognition and Neuroscience | 38 | Not Available |
| 11719 | Latin American Economic Review | 1 | Not Available |
| 11719 | Materials Horizons | 259 | Not Available |
| 11719 | Metallurgical Research & Technology | 2 | Not Available |
| 11719 | Oncology Research and Treatment | 11 | Not Available |
| 11719 | Physical Review Applied | 62 | Not Available |
| 11719 | Plastic Surgery | 1 | Not Available |
| 11719 | Progress in Tumor Research | 3 | Not Available |
| 11719 | Reviews in Fisheries Science & Aquaculture | 4 | Not Available |
| 11719 | Advanced Materials Interfaces | 161 | Not Available |

# Introduction

- The rationale for conducting the study should be clear from its introduction
- The introduction should provide a thorough review of the literature and identify gaps that the study will address
- Favorable and unfavorable findings about the drugs should be included
- The benefits and the risks associated with the treatment should be assessed
- The objective of the study should be clearly stated at the end of the introduction (in most studies)
- The hypotheses tested and the results expected can also be stated

# Type of hypotheses

Study hypotheses are tested statistically

- Null hypothesis: there is no difference between treatments or comparisons
- Alternative hypothesis: a difference is expected between therapies
  - One-tailed
    - An expected direction of the effect is stated
  - Two-tailed (Mostly used)
    - A change is expected but it can be in either direction

# Evaluating methods used

- Importance of a study's eligibility criteria, methods used for enrolling patients in a study (sampling) and informed consent
- Advantages and disadvantages of different controlled experimental design and types of control
- Importance of random assignment in a study
- Effect of adherance on study findings
- Importance of outcome measure selection
- Concepts of validity, reliability, sensetivity and specificity and their importance to the outcome measures used
- Dependent vs independent variables
- The levels of measurements

- Considerations in examining methods used in a study
  - Study sample
  - Sample size
  - Controlled experimental designs
  - Assignment to treatment groups
  - Blinding
  - Drug treatments
  - Adherence
  - Outcomes
  - Variables
  - Measurements

- It is almost impossible to design the perfect study
- One should differentiate between weaknesses and limitations that could invalidate the findings or just limit the applications of the results

# Eligibility (inclusion and exclusion) criteria

- Used to define the characteristics of the subjects enrolled in the study
- Inclusion criteria: characteristics that should be present in the subjects
- Exclusion criteria: characteristics that prevent subjects from participating  in the study
- Eligibility criteria defines the population for which the study results can be applied
- Selection bias occur when the study *sample* is chosen in a way that does not represent the target population

- Examples on exclusion criteria:
  - Nonstudy concurrent medications that might interact with the study drugs or have actions that affect the studied condition
  - Patients who have contraindications to the study drugs such as allergy or renal impairment
  - Patients who have other medical conditions that can interfere with the study findings

# Sampling (enrollment) considerations

- For best study sample, everyone in the population should have the same chance of being selected for the study
- Random methods are the best methods for sampling
- Random sampling is not always possible, since investigators can not have access to everyone in the population

# Types of sampling

- Simple random
- Stratified random
- Cluster
- Systematic
- Convenience

- Simple random sampling:
  - Everyone in the population is identified and a random procedure is used to identify persons for study inclusion
  - Example: computer generated sampling

- Stratified random sampling:
  - Enrolling similar numbers of patients who have or do not have certain characteristics (smokers or nonsmokers, diabetic or non-diabetic patients, etc.)
  - The population is divided into groups based on the presence or absence of a characteristic, and a random sample is chosen from each group for enrollment

- Cluster sampling:
  - All individuals present in identified clusters in the population are selected for enrollment
  - Example: everyone living in a city, or attending a hospital
- Systematic sampling:
  - Type of random sampling
  - Everyone in the population is known and the starting point is randomly selected
  - Selecting every nth person for the study

- Convenience sampling
  - Most commonly used in experimental studies
  - Nonrandom sampling that enrolls patients based upon advertisements or whether they are treated in a certain clinic that the investigators work at
  - Used when it is not possible to contact all persons in the target population
  - This is acceptable as long as it is defined at the time of the study enrollment which group the patient will be assigned to

Example:

- Investigators wish to study the efficacy of a new drug to increase smoking cessation, a condition in which subjects motivation to quit is very important.
- Patients are enrolled who respond to a newspaper ad asking for volunteers who would like to participate in a study to quit smoking
- Could selection bias be a problem?

# Informed consent

- Investigators need to ensure that the subjects in their studies are protected from harm to the extent possible
- Institutional review board (IRB) and the informed consent are used to ensure subjects are protected
- IRB is responsible for assuring that the subjects rights and welfare are protected, before and throughout the study
- The informed consent should be obtained by the investigators prior to enrollment in the study

# Informed consent

- Parts of informed consent include:
  - Providing a subject with adequate information about the study, and its benefits and risks
  - Giving the subject appropriate opportunity to consider all options
  - Responding to the subject's questions
  - Ensuring that the subject understands the information
  - Obtaining the subject's written voluntary consent to participate in the study
  - Providing additional information as needed

- With informed consent, the subjects also have the right to quit the study whenever they wish.

- The patients should also be aware of all the potential adverse effects and risks from each type of therapy they may receive

# Sampling size

- A study should have enough sampling size to identify a statistically significant difference among treatments when an effect exists

- Should be able to reject the null hypothesis

- With informed consent, the subjects also have the right to quit the study whenever they wish.

- The patients should also be aware of all the potential adverse effects and risks from each type of therapy they may receive

# Sample size

- The study should have enough sample size to identify a statistically significant difference among treatments
  - When the null hypothesis is rejected
- The extent to which a statistical test is able to identify a significant difference when there is an actual treatment effect is referred to as **power**
- Sample size is a key factor affecting a study's power
  - All other factors being equal, if the sample size increases the statistical power increases
- If a study's power is too low, the analysis could find the difference between treatments not to be statistically significant

# Sample size

- The study's sample size or the number of patient to enroll should ideally be calculated before the study begins
- The desired power is selected (≥80)
- The sample size can be calculated based on the power value

# Blinding

- Blinding (masking) is when patients or investigators do not know the intervention group that the patient was assigned to
- Reduces risk of bias
- Many outcome measures can be affected by personal believe that a therapy will work
  - Pain relief
  - Changes in mood
  - Development of side effects

# Blinding

- Single-blind: patients are unaware of the therapy they are receiving but the investigators know
- Double-blind: neither patients nor the investigators know which therapy the patients are receiving
- Triple-blind: If any non-investigators perform the analysis, neither them, patients, nor the investigator know which therapy the patients is receiving

- Double blinded studies are preferred to single blinded
  - The gold standard for RCT study designs
  - Less bias than single blind studies
- Sometimes blinding is harder (characteristic odor, taste, side effect)
- Unbinding occurs if a patient or investigator can identify what the patient is receiving during a blinded study
- How to tell if unblinding occurred?

# Outcomes

- A study's objective should specify the broad overall outcome of interest
  - Example: hypertension control, smoking cessation, diabetic control.
- The methods should clearly state the primary outcomes of interest as well as any secondary outcomes.
- The end point where the outcome will be successfully met should be specified when possible

# Variables

- Variables refer to study characteristics that can assume different values
- Dependent or independent variables, confounding variables
- Independent (explanatory) variables affect the value of the dependent (response) variables
  - Type of treatment
- Dependent variables: change in value as a result of the independent variable
  - The outcome measures are the dependent variables that can be altered by exposure to treatment

# Confounding variable

- A factor that can affect the value of the outcome measures in addition to the therapy being studied
- Affect the study results and their interpretation
- The confounding variables should be taken into account

# Measurements

- The tests or procedures used to measure changes in the desired outcomes (dependent variables) should be appropriate to the intended objectives

- Studies should select the best test or a combination of tests to measure outcomes

# Internal & external validity

- Internal validity: the extent to which a study's findings were appropriate and correct
  - The relationship between the intervention and outcomes was accurate
- External validity: The extent to which the findings of a study can be applied to the patients and settings  outside the study

- The stronger the study's design, methods and analyses, the greater the internal validity
  - RCT have greater internal validity than other study types
- External validity is important for applying the results from a study in clinical practice
- The study should provide clear protocols and definitions to ensure that an outcome measure is appropriately used throughout the study

# Scales of measurements

- Nominal
- Ordinal
- Continuous
  - Interval
  - Ratio

Determine the statistical tests used

# Scales of measurements

- Nominal (categorical): data that lack numerical qualities

  e.g. race, gender, presence or absence of adverse effect/ cure

- Ordinal: data that can be ranked on a scale with one value more or less than another, assigned numbers do not have exact differences

  e.g. opinions ranked using a scale of 5-1 (strongly agree-disagree)

  Ranking of severity of illness

# Scales of measurements

- Continuous: data that can assume an unlimited number of numerical values within a range with equal distances between numbers
  - Interval: Lack a true zero point
    - pH values, Fahrenheit or Celius temperatures
  - Ratio: have a true zero point
    - The most common type of continuous data used in clinical research
    - e.g. height, weight, drug concentration

# Examples

- Label as nominal, ordinal or continuous
  - Thyroxine serum concentration following thyroid replacement
  - The severity of neuropathic pain (3=severe, 2=moderate, 1= mild, 0= absent) following therapy with gabapentin or placebo
  - Number of osteoporosis patients who experienced a fracture during treatment with either alendronate (3 of 29, 10.3%) or risidronate (5 of 35, 14.3%)
  - Hemoglobin A1c concentration at baseline and following metformin therapy

# Statistical analyses

- Interpretation of a clinical trial's findings usually depends upon statistical analyses of the data

- Statistics help the investigators and readers find out:
  - If differences found in outcome measures resulted from treatments or not
  - The association among variables measured in the study group
  - Predictions that can be made for the populations based on the results from the study sample

- Many statistical tests are available
  - The appropriate test is chosen depending on the types of data and conditions involved

# One-tailed vs two-tailed tests

- The use of one-tailed or two-tailed tests depends on the study's stated objectives or hypothesis
- One-tailed test is used only when a study clearly states a one-tailed hypothesis (unidirectional change)
- Most clinical studies use two-tailed tests

- Factors considered when selecting a statistical test:
  - Level/scale of measurement of the data being analyzed (nominal, ordinal, continuous)
  - Number of treatment groups being compared
  - Data collected from paired (the same patient) or unpaired (different) patients

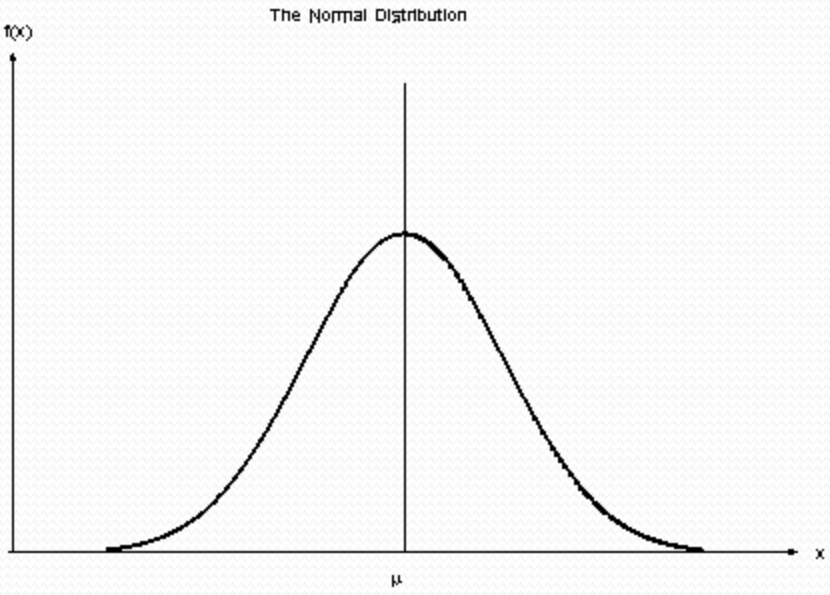
- Categories of statistical tests
  - Parametric
  - Nonparametic

- Choice of parametric or nonparametric tests to analyze data depends on the population from which the study sample was selected

# Parametric tests

- Used when the data being analyzed is continuous and normally (or near normally) distributed
- Normal (Gaussian)distribution resembles a bell-shaped curve when graphed by frequency
- Preferred

  (more statistical power)
- Example:
  - *t-test*
  - ANOVA



# Parametric tests

- Continuous data
- Normally distributed data
- Population variances are equal (or nearly equal)
- Observations or measurements within a population are independent

# Parametric tests

- *t-test*
  - Used when comparing the means of only 2 groups
  - Paired or unpaired *t-test*
- ANOVA
  - Used when comparing the means of 3 or more groups
  - If ANOVA is significant, a multiple comparison (post-hoc) test is used to identify which 2 group mean comparison is statistically significant
    - Scheffe's test
    - Tukey honestly significant (HSD) test
    - Dunnett test
    - Fisher least significant test (LSD)

- ANOVA
  - One-way ANOVA
    - ≥3 groups, **one** independent variable, parallel study design (unpaired)
  - Two way ANOVA
    - ≥3 groups, **two** independent variable, parallel study design (unpaired)
  - Repeated measures ANOVA
    - ≥3 groups, **one** independent variable, cross-over design, **(paired data)**
    - ≥2 groups, **one** independent variable, parallel study design, **multiple measurements** taken over time in each study group

# Example

- A parallel double-blind study is performed to compare the diastolic blood pressure after 12 weeks of therapy in patients randomized to receive enalapril ($n$ = 68), lisinopril ($n$ = 72), fosinopril ($n$ = 65)
- Assume the blood pressure readings are normally distributed

- Which statistical test should be used to analyze the results?
  A. Paired t-test
  B. Unpaired t-test
  C. One-way ANOVA
  D. Two-way ANOVA
  E. Repeated measures ANOVA

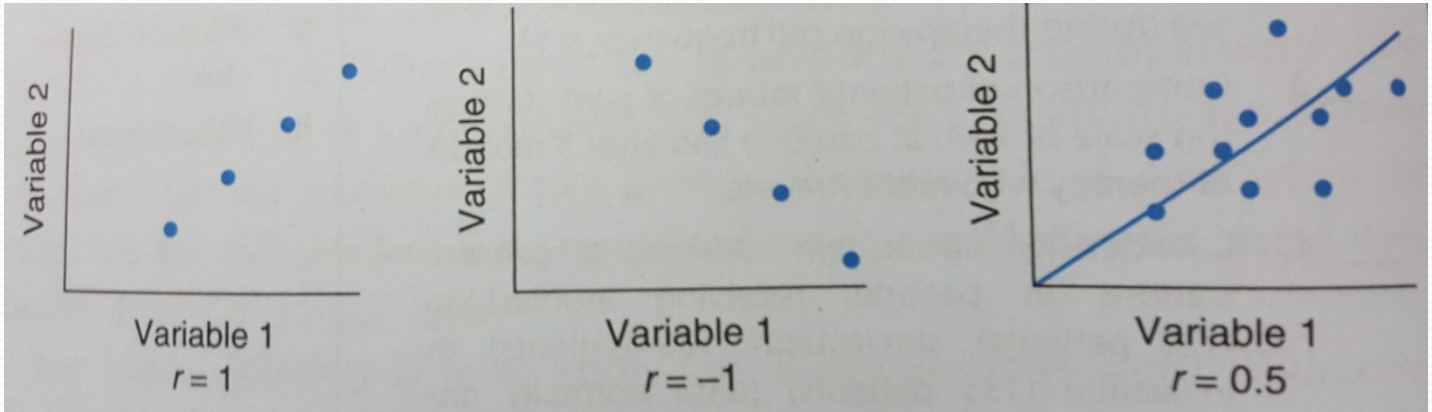
# Nonparametric test

- Used when data is not normally distributed

- The choice of nonparametric test depends on:
  - Whether the data being analyzed are nominal or ordinal
    - (*Chi-square*, *Fisher's exact* are examples on tests used for nominal data)
    - (*Mann-Whitney*, *Friedman* are examples on tests used for ordinal data)
  - The number of groups involved
  - Samples are paired or unpaired

# Correlation

- The association between two variables
- Correlation coefficient (*r*) is used to quantify the degree and direction of a linear association between 2 variables
  - *r* ranges from -1 to 1
    - Negative *r* indicates an inverse association
    - Positive *r* indicates a positive association
    - *r* = 0; no association



# Correlation

# Evaluating results, interpretation and conclusions of clinical studies

- Interpreting a study's data and significance of the results are important for applying findings to clinical practice or not
  - Measures of central tendency
  - Measures of variability
  - Hypothesis testing
  - Statistical inference
  - Conclusions
  - Application to clinical practice

# Measures of central tendency

- Central tendency of the data reflects the usual or typical response to therapy in the study

- 3 Central tendency measures:
  - The mean
  - The median
  - The mode

# The mean

- Provides a good estimate for central tendency (clustering) of continuous data
- Can be used for ordinal-level data
  - Keep in mind that distances between numbers are not equal
  - Caution should be taken when interpreting ordinal data represented by the mean
- In the presence of outliers, the mean can misrepresent the data

# Examples

- Serum potassium values in mEq/L for 10 patients are: 4.1, 3.1, 5.2, 3.7, 5.1, 3.2, 4.8, 4.3, 3.9, 5.1
  - The mean = Sum/n= 4.25

- Estrogen concentrations in 10 women in pg/mL

  28, 29, 30, 30, 29, 28, 28, 30, 30, 259
  - The mean = 52.1
  - Without the outlier, the data is clustered around 29

# The median

- The midpoint of a rank ordered data
- The 50th percentile
- Can be used for ordinal or continuous data
- It better represents the central tendency of data with one or more outliers

# Examples

- Estrogen concentrations in 10 women in pg/mL

  28, 29, 30, 30, 29, 28, 28, 30, 30, 259

  - The mean = 52.1
  - The median = 29

# The mode

- The most frequently occurring value in a data set
- Can be used for nominal, ordinal, or continuous data
  - Only one used for nominal data
  - Not very helpful for continuous data

# Examples

- Systolic blood pressure in patients in mmHg

  <120 (23 patients)

  120-149 (27 patients)

  ≥150 mmHg (18 patients)

- Patients satisfaction with therapy on a scale of 0-4

  If most patients indicated 3, the mode =3

# Example

- A study evaluating the efficacy of herbal Chinese tea extract for treating hyperlipidemia in 12 diabetic patients reported that serum cholesterol levels following 8 weeks of therapy were:

  **mean = 220** mg/dL; **median =175** mg/dL

- Which value appears to provide a better estimate of the central tendency of the data?

# Measures of variability

- Measures of spread or dispersion of the data
  - The range
  - Interquartile range (IQR)
  - Variance (not used frequently)
  - Standard deviation (SD)
    - Most commonly used
    - Reported as mean ± SD or mean (SD)
  - Standard error of the mean (SE or SEM) =SD/
    - Frequently used in clinical trials but should not be used

# Example

- Two studies examined whether counseling diabetes patients about their medications increased blood glucose control.
- The patients' mean (SD) fasting blood glucose concentration following counseling were

Study 1 160 (31) mg%

Study 2 158 (45) mg%

- Which study reported greater variability in individual patient responses following therapy?

# Example

- A study examined the efficacy of a new drug for hypertension treatment in 200 patients. At the end of the study (week 16), the change from baseline in mean (SEM) systolic/diastolic blood pressure was **-18.07(0.8)/-10.9(0.5)**
- Is it appropriate for SEM to be reported

# *P* values

- *P* values provides the likelihood that chance was responsible for the effect observed or that the null hypothesis was true

- *P* values range from zero to 1

- The value of 0.05 (level of significance) is used as a cut-off in clinical studies
  - $P < 0.05$ statistically significant findings
  - $P \geq 0.05$ findings not statistically significant

# Example

- A study compared the efficacy of oral mesalazine (*n*=28 patients) with topical mesalazine (*n*=30 patients) for the treatment of distal ulcerative colitis.
- Following 2 weeks of therapy with either agent the clinical response ate was 43% with oral mesalazine vs 58% with topical mesalazine (*P*=0.003)
- Is therapy difference statistically significant?

# Statistical significance vs clinical significance

- If a study is statistically significant ($p<0.05$), we should then look at the treatment effect or difference between group to determine clinical significance
- Can a study be clinically significant if it is not statistically significant?

# Discussion section

- Summarizes all important findings of the study
- Analyzes the results in relation to previous studies or other relevant literature
- Explanation of the results can be provided
- Limitations of the study should be mentioned
- Types of future research needed in the area
- A final summary that states the study's conclusions and clinical applicability of the findings

# Evaluating clinical studies

- Type of study (strength and limitations)
- Introduction and study rationale
- Enrollment of subjects
- Treatment regiments
- Outcome measures
- Data handling and statistical analyses
- Presentation and interpretation of the results
- Author's discussion and conclusions

# Key questions when critiquing published experimental studies

- Journals and authors
  - Does the journal have an editorial board? Does the journal use peer review?
  - Any potential conflict of interest for authors or investigators? Would they affect the objective, methods or conclusions?
- Introduction
  - Was appropriate scientific background or rationale provided?
  - Is the stated objective or hypothesis consistent with the research question needed to be addressed?
  - **Is the study adequately designed to fulfill its stated objective?**

# Key questions when critiquing published experimental studies

- Patients/subjects
  - Were the inclusion and exclusion criteria appropriate and representative of the population of interest?
  - Were factors that might interfere with the study excluded?
  - Was the number of patients enrolled and analyzed sufficient to maintain at least 80% power for outcome measures?

# Key questions when critiquing published experimental studies

- Treatment regimens
  - Appropriate control was used?
  - Was dosing and administration representative to what would be used in practice?
  - Was a concurrent control design used? If not were sufficient wash-out periods used? Was carry-over effect analyzed?
  - **Did the study randomly assign patients into groups?**
  - **Was the study blinded?** Was unblinding a problem?
  - **Were the drugs administered for a sufficient duration?**
  - If concurrent medications were allowed, was their use similar among groups?
  - Were adverse effects reported and statistically analyzed?
  - Was adherence to treatments and study requirements measured?

# Key questions when critiquing published experimental studies

- Outcome measures
  - **Were the primary and secondary outcome measures clearly defined and appropriate for the objective?**
  - Were standardized methods used?
  - Was the timing of outcome measures appropriate and of adequate frequency?
  - Were different patient groups handled similarly?
- Statistical methods
  - Were appropriate statistical tests used for all outcome measures?
  - Did any reported correlation (r) values represent strong or clinically important associations?

# Key questions when critiquing published experimental studies

- Results
  - Were any significant differences apparent among groups?
  - Was the number of patients accounted for at each step? And was it clear how many patients were in each analysis?
  - Drop-outs and data handling method used?
  - Was power appropriate for outcome analysis?
  - Was the measure of central tendency appropriate?
  - Were the measures of variability appropriate and sufficient?
  - Were findings statistically significant? And if yes, were they large enough to be clinically significant?

# Key questions when critiquing published experimental studies

- Discussion
  - Were the results (positive and negative) interpreted?
  - Did the authors adequately explain key study limitations and any discrepancies from other similar studies?
  - Were conclusions consistent with the results and study limitations?
- Overall assessment
  - What were the important weaknesses of the study? What key findings should be taken away from the study?
  - Could any study limitations or design weaknesses reduce internal validity thereby affecting its external validity?
  - What is the role of the study in clinical practice?
  - Is any further research needed?