# Epidemiological Study Designs (Case Control, Cohort, RCT)

### 1. Observational Studies

In observational studies, nature is allowed to take its own course; the investigator measures but does not actively intervene [2]. They are subdivided into:

### A. Descriptive Studies

- These are usually the first phase of an epidemiological investigation and are concerned with observing the distribution of disease in a population by time, place, and person [3] [4].
- **Primary Purpose:** Formulation of an aetiological hypothesis [5].

### B. Analytical Studies

- These studies go further by analyzing the relationship between health status and other variables to test the hypotheses formulated in descriptive studies [2] [5]. They are classified into four distinct types [1]:
    1. **Ecological or Correlational studies:** The unit of study is *populations*.
    2. **Cross-sectional or Prevalence studies:** The unit of study is *individuals*.
    3. **Case-control or Case-reference studies:** The unit of study is *individuals*.
    4. **Cohort or Follow-up studies:** The unit of study is *individuals*.

## 2. Experimental Studies (Intervention Studies)

Experimental studies involve an active attempt by the investigator to change a disease determinant or alter the progress of a disease [2].

- **Primary Purpose:** Confirmation of the aetiological hypothesis [5].
- They are classified into three types based on their unit of study [1]:
    1. **Randomized Controlled Trials (or Clinical trials):** The unit of study is *patients*.
    2. **Field trials:** The unit of study is *healthy people*.
    3. **Community trials (or Community intervention studies):** The unit of study is *communities*.

# I. Master Case-Control Questions

- **The Ultimate Scenario:** A surgeon observes a high number of rare oral mucosa carcinoma cases linked to chewing specific plant leaves. Concurrently, an Anganwadi observes high malnutrition, matching 50 underweight children with 50 ideal-weight controls.
    - Identify the ideal study design for these scenarios.
    - Describe the exact steps to conduct this study (including matching).
    - How do you calculate the measure of risk/association ($OR = \frac{a \times d}{b \times c}$)?
    - What are the specific biases (e.g., recall bias, selection bias) and confounding factors involved, and how do you overcome them?
    - List the advantages and disadvantages of this design.

**Scenario 1: Rare Oral Mucosa Carcinoma**

- **Ideal Study Design:** Case Control Study.
- **Reasoning:** Case control studies are particularly suitable and recommended for investigating rare diseases or conditions about which little is known [1]. Conversely, cohort studies are inappropriate when the disease under investigation is rare because they would require an unfeasibly large number of subjects and a long follow-up period to yield results [2] [3]. By using a case control design, the surgeon can start directly with the rare carcinoma cases and work backwards to test the hypothesis regarding the suspected exposure (chewing specific plant leaves) [4].

**Scenario 2: Malnutrition with Matched Controls**

- **Ideal Study Design:** Case Control Study.
- **Reasoning:** This scenario perfectly describes the framework and basic steps of a case control study, which starts by identifying a group of "cases" (the 50 underweight children) and a comparison group of "controls" (the 50 ideal-weight children) [4] [5]. Furthermore, pairing the cases with controls represents "matching," a core procedural step in case control studies used to eliminate confounding factors and ensure comparability between the two groups [6] [7].

There are four basic steps involved in conducting a case control study [1].

## 1. Selection of Cases and Controls

- **Selection of Cases:** This requires a crucial prior definition of what constitutes a "case," which includes specifying exact diagnostic criteria and eligibility criteria (such as requiring that only newly diagnosed, or "incident," cases within a specified period are eligible) [2], [3]. Cases can be drawn from either a single hospital, a network of hospitals, or the general population [4].
- **Selection of Controls:** Controls must be free from the disease under study and must be as similar to the cases as possible [5]. They can be selected from hospitals (patients with different illnesses), relatives (spouses and siblings), neighbours, or the general population [6], [7]. While one control per case is standard for large studies, small studies (under 50 subjects) may use 2, 3, or even 4 controls for each study subject [8].

## 2. Matching

- **Definition and Purpose:** Matching is the process by which controls are selected in such a way that they are similar to cases with regard to certain pertinent selected variables (e.g., age, sex, occupation) [9] . The primary purpose of matching is to ensure comparability and neutralize "confounding factors"—variables that are associated with both the exposure and the disease, and could distort the results if distributed unequally [9] , [10] .
- **Crucial Precaution:** The suspected aetiological factor or the variable you wish to measure must never be matched, because doing so would automatically make the cases and controls alike with respect to that factor, thereby eliminating its aetiological role from the study [11] .
- **Procedures:**
    - *Group matching:* Cases are assigned to sub-categories (strata) based on characteristics, and appropriate controls are established to ensure the frequency distribution is similar in both groups [12] .
    - *Matching by pairs:* For each individual case, a control is chosen which can be matched quite closely (e.g., finding a 50-year-old mason without the disease to match a 50-year-old mason with the disease) [12] .

## 3. Measurement of Exposure

- Definitions and criteria regarding exposure must be established, and information must be obtained in precisely the same manner for both cases and controls [13].
- Data can be gathered through interviews, questionnaires, or by studying past records, such as hospital or employment records [13].
- It is vital during this stage to consider and rule out "bias," which is any systematic error in the determination of the association between the exposure and disease [13], [14].

**4. Analysis and Interpretation** The final step aims to discover exposure rates and estimate the disease risk [15].

- **Exposure Rates:** This is a direct estimation of the frequency of exposure in the disease group versus the non-disease group [15]. Statistical tests, such as the standard error of difference between two proportions or the Chi-square test, are used to ascertain if a statistical association exists between the exposure status and the occurrence of the disease [16], [17].
- **Estimation of Risk:** Because a case control study does not provide incidence rates (due to the lack of an appropriate denominator or population at risk), the relative risk cannot be calculated directly [18], [19]. Instead, the risk is estimated using the **Odds Ratio** (Cross-product ratio), calculated as $ad/bc$ [19], [14]. The Odds Ratio measures the strength of the association between the risk factor and the outcome [19].

**1. Constructing the 2x2 Contingency Table** The variables $a$, $b$, $c$, and $d$ in the formula represent the following distinct groups:

- **a**: Cases (Disease present) who were exposed to the suspected risk factor [1] [2].
- **b**: Controls (Disease absent) who were exposed to the suspected risk factor [1] [2].
- **c**: Cases (Disease present) who were NOT exposed to the suspected risk factor [1] [2].
- **d**: Controls (Disease absent) who were NOT exposed to the suspected risk factor [1] [2].

**2. Calculation of the Odds Ratio** Because a case control study does not provide exact incidence rates (due to the lack of an appropriate population-at-risk denominator), the exact Relative Risk cannot be calculated directly [3] [4]. Instead, the Odds Ratio is calculated as the cross-product of the table entries to estimate the relative risk [2] [4]:

$$OddsRatio = \frac{a \times d}{b \times c}$$

**3. Essential Assumptions** The derivation of the odds ratio to estimate relative risk is based on three key assumptions [4]:

1. The disease being investigated must be relatively rare.
2. The cases must be representative of all those with the disease.
3. The controls must be representative of all those without the disease.

**4. Interpretation of the Results** The resulting OR value measures the strength of the association between the risk factor and the outcome [2] [4]:

- **OR = 1:** Indicates that there is *no association* between the exposure and the outcome [2].
- **OR > 1:** Indicates a *positive association* between the exposure and the outcome (i.e., the exposure increases the risk of developing the disease) [2].
- **OR < 1:** Indicates a *negative association* (i.e., the exposure is actually a protective factor against the disease) [2].

## 1. Specific Biases in Case Control Studies

- **Memory or Recall Bias:**
  - *Concept:* When asked about past history, "cases" (those who have the disease) are more likely to recall certain events, habits, or exposures compared to healthy "controls" [2].
  - *How to overcome:* While difficult to eliminate completely since the study relies on memory or past records [3], obtaining information about exposure in precisely the same manner for both cases and controls (e.g., using objective past records like hospital or employment records instead of relying solely on memory) can help mitigate this [4].
- **Selection Bias:**
  - *Concept:* This occurs when the selected cases and controls are not truly representative of the cases and controls in the general population, leading to systematic differences in their characteristics [2].
  - *How to overcome:* Selection bias is best controlled by preventing it during the study design phase [2]. Epidemiologists recommend selecting cases from one source and obtaining controls from more than one source to avoid the influence of this bias [5].

- **Berkesonian Bias:**
  - *Concept:* A specific type of selection bias recognized by Dr. Joseph Berkeson. It arises from the different rates of admission to hospitals for people with different diseases, skewing results when both cases and controls are selected exclusively from a hospital setting [2] [6].
- **Interviewer's Bias:**
  - *Concept:* If the interviewer knows the study's hypothesis and knows who the "cases" are, they may subconsciously question the cases more thoroughly than the controls regarding the suspected causal factor [6].
  - *How to overcome:* This bias can be entirely eliminated by **double-blinding**, ensuring neither the interviewer nor the participant knows the group allocations [6]. A useful check is also to record and compare the average length of time taken to interview cases versus controls [6].

## 2. Confounding Factors

- **Concept:** A confounding factor is a variable that is associated with both the exposure and the disease, and is distributed unequally between the study and control groups [7] [8]. It acts as an independent risk factor for the disease [8]. If not adequately accounted for, it can distort or completely confound the results [7]. For example, in a study of alcohol consumption (exposure) and oesophageal cancer (disease), smoking is a confounding factor because it is linked to alcohol use and is an independent cause of oesophageal cancer [8].
- **How to overcome:**
    - **Matching:** The primary method to remove confounding bias in case control studies is matching [2] [7]. This involves selecting controls so that they are similar to the cases regarding pertinent confounding variables (like age, sex, or occupation) [7]. This can be done via *group matching* (stratification) or *matching by pairs* [9].
    - **Other Methods:** Confounding can also be controlled through randomization, restriction, stratification, and statistical modelling [10].

## ADVANTAGES

- **Ease and Cost:** They are relatively easy to carry out, rapid, and inexpensive, especially when compared to cohort studies [1] [2].
- **Sample Size:** They require comparatively few subjects to conduct the study [1] [2].
- **Suitability for Rare Diseases:** They are particularly suitable for investigating rare diseases or conditions about which very little is known [1] [2].
- **Safety and Ethics:** There is no risk to the subjects involved, ensuring that ethical problems remain minimal [1] [2].
- **Multiple Exposures:** The design allows for the study of several different aetiological factors (multiple exposures) that may result in a single disease outcome [2] [3].
- **Practical Application:** By identifying risk factors, rational prevention and control programmes can be established [3].
- **No Attrition:** There are no attrition (loss to follow-up) problems because the study does not require tracking individuals into the future [2] [3].

## DISADVANTAGES

- **High Risk of Bias:** They are highly susceptible to biases, particularly recall bias and selection bias, because the accuracy of data relies on memory or past records, making validation difficult or impossible [3] [4].
- **Control Group Selection:** Selecting an appropriate and comparable control group can be difficult [5].
- **Inability to Measure Incidence:** They cannot measure true disease incidence; they can only estimate the relative risk by calculating an odds ratio, which is an inferior measure of the strength of association compared to actual relative risk [4] [5].
- **Causation vs. Association:** They cannot distinguish between direct causes and merely associated factors, and the temporality of the association cannot be established [4] [5].
- **Natural History:** The natural history of the disease cannot be studied using a case control design [4].
- **Treatment Evaluation:** They are not suited for evaluating the efficacy of therapy or prophylaxis of a disease [5].
- **Representativeness:** A major ongoing concern is whether the selected cases and controls are truly representative of the general population [5].

## II. Master Cohort Questions

- **The Ultimate Scenario:** 700 out of 5500 babies weighed below 2500g at birth. After a 1-year follow-up, 500 total infant deaths occurred, 350 of which were normal birth weight. Simultaneously, a medical officer investigates if cotton mill workers' health is affected by their work environment.
  - What type of study design is this?
  - Discuss the framework and steps for conducting this study.
  - Construct a 2x2 table based on the infant mortality data provided.
  - How do you analyze the data? (Explain the calculation of Relative Risk and Attributable Risk).
  - Mention the merits and demerits of this study design.

## Scenario 1: Low Birth Weight and Infant Mortality

- **Study Design:** Cohort Study (Specifically, a Prospective Cohort Study).
- **Reasoning:** The distinguishing feature of a cohort study is that it proceeds forward from cause to effect [1]. In this scenario, the study begins by identifying a cohort of 5500 newborns and categorizes them by their exposure status: an exposed group (the 700 babies with low birth weight) and a non-exposed control group (the remaining babies with normal birth weight) [2]. These defined study groups are then followed up over a period of time (1 year) to observe and determine the outcome or frequency of the event (infant deaths) [1] [3]. Because the disease or outcome had not yet occurred at the time the investigation began, it perfectly fits the prospective cohort design [4].

## Scenario 2: Cotton Mill Workers' Health

- **Study Design:** Cohort Study (Exposure Cohort).
- **Reasoning:** In epidemiology, when an exposure is rare or highly specific, it is more economical and practical to select a cohort of persons already known to have experienced the exposure [5]. A readily accessible source for these "special groups" or "exposure cohorts" are workers employed in specific industries [5]. By selecting cotton mill workers, the medical officer is assembling a cohort based on a shared environmental exposure (e.g., cotton fibre dust, which is known to cause byssinosis) [1] [6]. The officer can then follow this group prospectively or retrospectively to evaluate the subsequent development of respiratory or other health outcomes, moving from the suspected cause to the effect [1].

The basic framework of a cohort study works from "cause to effect" [1]. It begins with a group, or cohort, exposed to a suspected aetiological factor (the "study cohort") and a group not exposed to that factor (the "control cohort") [2]. Before the study begins, the exposure has occurred, but the disease has not [1]. Both cohorts are then followed forward under identical conditions over a period of time to determine the outcome of exposure (e.g., onset of disease, disability, or death) [3].

The exact steps (or elements) involved in conducting a cohort study are:

**1. Selection of study subjects** Subjects can be assembled in one of two ways:

- **General population:** Cohorts may be assembled from individuals residing in well-defined geographical, political, and administrative areas (e.g., the Framingham Heart Study) [4].
- **Special groups:** These include professional groups (e.g., doctors, nurses, teachers), insured persons, or specific "exposure groups" (e.g., radiologists or industrial workers who are known to have experienced a specific exposure) [5] [6].

**2. Obtaining data on exposure** Information regarding exposure levels must be obtained to classify cohort members. Sources of data include:

- **Cohort members:** Through personal interviews or mailed questionnaires [7].
- **Review of records:** Utilizing past hospital, employment, or medical treatment records [7] [8].
- **Medical examinations or special tests:** Such as blood pressure, serum cholesterol, or ECG readings [8].
- **Environmental surveys:** To obtain information on exposure levels in the environment where the cohort lived or worked [8].

**3. Selection of comparison groups** To evaluate the experience of the exposed group, an appropriate comparison group is required:

- **Internal comparisons:** A single cohort enters the study and its members are classified into sub-groups based on varying degrees or levels of exposure to risk (e.g., amount of cigarettes smoked) [9].
- **External comparisons:** Putting up a separate external control group that is demographically similar but lacks the exposure (e.g., comparing a cohort of radiologists with a cohort of ophthalmologists) [10].
- **Comparison with general population rates:** Used when no specific control group is available. The mortality or morbidity experience of the exposed group is compared against the general population rates in the same geographic area [11].

**4. Follow-up** Regular follow-up of all participants is necessary to assess the outcome. Procedures include:

- Periodic medical examinations of each cohort member [12].
- Reviewing physician and hospital records [12].
- Routine surveillance of death records [12].
- Mailed questionnaires, telephone calls, or periodic home visits [12].
- *Note:* Some percentage of losses to follow-up (attrition) is inevitable due to death, migration, or loss of interest, but every effort must be made to minimize this [12].

**5. Analysis** The data are analyzed to determine two key metrics:

- **Incidence rates:** Calculated directly for the outcome in both the exposed and non-exposed groups [13].
- **Estimation of risk:** This is done by calculating the **Relative Risk (RR)**, which is the ratio of the incidence of the disease among the exposed to the incidence among the non-exposed [14]. Furthermore, **Attributable Risk** can be calculated to indicate the extent to which the disease under study can be attributed to the specific exposure [14].

### 1. Construction of the 2x2 Contingency Table

To analyze the data, we first categorize the 5500 babies into a standard 2x2 contingency table based on their exposure status (Low birth weight vs. Normal birth weight) and the outcome (Death vs. Survival) [1] [2].

- **Total Cohort:** 5500 babies.
- **Exposed (Low Birth Weight):** 700 babies.
- **Non-Exposed (Normal Birth Weight):** $5500 - 700 = 4800$ babies.
- **Outcome (Total Deaths):** 500 infant deaths.
- **Deaths in Non-Exposed:** 350 deaths.
- **Deaths in Exposed:** $500 - 350 = 150$ deaths.

| Suspected Risk Factor | Developed Outcome (Infant Death) | Did Not Develop Outcome (Survived) | Total |
|---|---|---|---|
| **Exposed** (Low Birth Weight) | **150** (a) | **550** (b) | **700** (a+b) |
| **Not Exposed** (Normal Weight) | **350** (c) | **4450** (d) | **4800** (c+d) |

## 2. Analysis of the Data

The data in a cohort study are analyzed by calculating the incidence rates of the outcome among the exposed and non-exposed groups, followed by the estimation of risk using Relative Risk and Attributable Risk [2] [3].

### Step A: Calculate Incidence Rates

- **Incidence among exposed ($I_{exp}$):** $\frac{a}{a+b} = \frac{150}{700} = 0.2143$ (or $214.3$ deaths per 1000) [2].
- **Incidence among non-exposed ($I_{non-exp}$):** $\frac{c}{c+d} = \frac{350}{4800} = 0.0729$ (or $72.9$ deaths per 1000) [2].

### Step B: Calculate Relative Risk (RR)

- **Definition:** Relative risk is the ratio of the incidence of the disease (or death) among exposed to the incidence among non-exposed [3]. It is a direct measure (or index) of the "strength" of the association between the suspected cause and effect [4].
- **Formula:** $RR = \frac{\text{Incidence among exposed}}{\text{Incidence among non-exposed}}$ [3].
- **Calculation:** $RR = \frac{214.3}{72.9} = \mathbf{2.94}$
- **Interpretation:** A relative risk greater than 1 suggests a "positive" association [4]. In this scenario, babies with low birth weight are **2.94 times more likely** to die in their first year compared to babies with normal birth weight.

## Step C: Calculate Attributable Risk (AR)

- **Definition:** Attributable risk is the difference in incidence rates of disease (or death) between an exposed group and non-exposed group, often expressed as a percentage [5]. It indicates to what extent the outcome under study can be directly attributed to the exposure [6].
- **Formula:** $AR = \frac{\text{Incidence rate among exposed} - \text{Incidence rate among non-exposed}}{\text{Incidence rate among exposed}} \times 100$ [6].
- **Calculation:** $AR = \frac{214.3 - 72.9}{214.3} \times 100 = \frac{141.4}{214.3} \times 100 = \mathbf{65.98\%}$
- **Interpretation:** This figure indicates that if the association is causal, approximately **66% of the infant deaths among the low birth weight group were directly due to their low birth weight** [6]. This suggests the amount of disease (mortality) that might be eliminated if the factor under study (low birth weight) could be controlled or prevented [6].

## MERITS (ADVANTAGES)

- **Calculates Incidence:** Incidence rates can be calculated directly [1].
- **Multiple Outcomes:** Several possible outcomes related to a single exposure can be studied simultaneously (e.g., a cohort study on smoking can simultaneously study its association with lung cancer, coronary heart disease, peptic ulcer, etc.) [1].
- **Direct Estimate of Risk:** Provides a direct estimate of relative risk [1].
- **Dose-Response:** Dose-response ratios can be calculated [1].
- **Minimizes Bias:** Since comparison groups are formed before the disease develops, certain forms of bias (like misclassification of individuals into exposed and unexposed groups) can be minimized [1].

## DEMERITS (DISADVANTAGES)

- **Large Sample Size:** Cohort studies require a large number of people, making them difficult to carry out [2] [3].
- **Unsuitable for Rare Diseases:** They are generally unsuitable for investigating uncommon diseases or diseases with a low incidence in the population [2].
- **Long Duration:** It takes a long time to complete the study and obtain results (often 20–30 years). During this time, investigators may die, or participants may change their classification [2].
- **Administrative Problems:** Extensive record-keeping is inevitable, along with potential loss of funding and experienced staff over the years [3].
- **Attrition:** A substantial proportion of the original cohort may be lost due to migration, death, loss of interest, or refusal to provide information [3].

- **Selection Difficulties:** Finding comparison groups that are truly representative of the exposed and unexposed segments of the population is a limiting factor (e.g., volunteers may not be representative) [3].
- **Changes in Criteria:** There may be changes in standard diagnostic criteria or methods of the disease over the prolonged follow-up period [3].
- **High Cost:** They are highly expensive [4].
- **Behavioural Changes:** The study itself may alter people's behaviour (e.g., increased concern in a study about smoking may induce subjects to stop or decrease smoking) [4].
- **Ethical Problems:** As evidence accumulates about a disease's aetiological factor, investigators face the ethical obligation to intervene and potentially reduce or eliminate that factor [4].

## III. Master Randomized Control Trial (RCT) Questions

- **The Ultimate Scenario:** A drug company claims a new sweetened zinc beverage reduces diarrhea in children under five. A local practitioner also claims a specific capsule prevents thrombocytopenia in dengue patients.
  - What is the appropriate study design to test the efficacy of these interventions and why?
  - List and describe the basic steps in conducting this trial.
  - Explain the concepts and importance of Randomization and Blinding in this study.
  - List and briefly describe TWO ethical considerations (e.g., placebo use) that must be addressed, and explain why this design faces more ethical issues than observational studies.

- **Evaluation of New Therapies:** The RCT is considered the "No. 1 method of evaluation" for new therapies, drugs, or programmes [4]. Clinical trials are specifically concerned with evaluating the efficacy of therapeutic agents, making them the exact design needed to test the claims of the drug company and the local practitioner [2] [3].
- **Providing Scientific Proof:** Experimental studies like RCTs are designed to provide "scientific proof" and offer a reliable method for measuring the effectiveness and efficiency of treatments and preventive measures [1] [5].

- **Manipulation of Variables:** Unlike observational studies where the epidemiologist only observes the natural course of events, an RCT involves an active attempt to change a disease determinant or its progress [5] [6]. The investigator intervenes by deliberately applying the suspected causal factor (the zinc beverage or the capsule) to the experimental group while making no change to the control group, and then compares the outcomes [5] [...].
- **Elimination of Bias and Confounding:** RCTs utilize randomization to allocate participants into "study" and "control" groups [9]. This process is the "heart" of a control trial because it ensures comparability between groups, eliminates "selection bias," and removes both known and unknown confounding factors, providing the greatest confidence that the findings are valid [9] [...].

**1. Drawing up a protocol** One of the essential features of an RCT is that it must be conducted under a strict protocol [1]. The protocol specifies the aims and objectives, questions to be answered, criteria for selecting study and control groups, sample size, allocation procedures, treatments to be applied, and standardization of working procedures [1]. Once evolved, the protocol must be strictly adhered to throughout the study to prevent bias and reduce sources of error [1].

**2. Selecting reference and experimental populations**

- **Reference (or target) population:** This is the broader population to which the findings of the trial will be applicable if the intervention is found successful (e.g., a specific age group, occupational group, or mankind in general) [2].
- **Experimental (or study) population:** This is the actual population derived from the reference population that participates in the study [3]. To minimize losses to follow-up, a stable population is chosen, and volunteers must fulfill three criteria:
    1. They must give "informed consent" after being fully informed of the trial's purpose and dangers [4].
    2. They must be representative of the reference population [4].
    3. They must be qualified or eligible for the trial (i.e., fully susceptible to the disease under study) [5].

**3. Randomization** Randomization is a statistical procedure by which participants are allocated into "study" and "control" groups [6]. It is considered the "heart" of a control trial because it eliminates "selection bias" and allows for true comparability [6] [7]. Unlike matching, which only controls for known factors, randomization ensures that both known and unknown confounding factors are distributed equally between the two groups [7].

**4. Manipulation (or intervention)** Once groups are formed, the investigator intervenes by deliberately applying, withdrawing, or reducing the suspected causal factor (e.g., a drug or vaccine) in the study group [8]. This manipulation creates an independent variable, the effect of which is subsequently determined by measuring the final outcome (the dependent variable) [9].

**5. Follow-up** Both the experimental and control groups are examined at defined intervals of time, in a standard manner, with equal intensity, and under the same circumstances until the final assessment [9]. Some losses to follow-up are inevitable due to death, migration, or loss of interest—this is known as "attrition"—but every effort must be made to minimize it to generalize the results accurately [10].

**6. Assessment of outcome** The final step is evaluating the outcome of the trial in both groups. Assessment is done in terms of:

- **Positive results:** The benefits of the experimental measure, such as reduced incidence or severity of the disease [10].
- **Negative results:** The severity and frequency of adverse side-effects and complications [10].
- **Blinding:** To prevent systematic errors during assessment—such as "subject variation" (patients feeling subjectively better), "observer bias" (investigator influenced by knowing the treatment), and evaluation bias—a technique called "blinding" is used [11] [12]. Trials can be **Single Blind** (participant is unaware), **Double Blind** (neither doctor nor participant is aware), or **Triple Blind** (participant, investigator, and data analyzer are all blind to group allocation) [12] [13].

## RANDOMIZATION

**Concept:** Randomization is a statistical procedure by which the participants are allocated into groups, usually called "study" and "control" groups, to receive or not to receive an experimental preventive or therapeutic procedure, manoeuvre, or intervention [1]. By random allocation, every individual gets an equal chance of being allocated into either group [2]. It is best done by using a table of random numbers [3]. Crucially, randomization is done only after the participant has entered the study, qualified for the trial, and given their informed consent [3].

**Importance:**

- **The "Heart" of the Trial:** Randomization is the "heart" of a control trial because it gives the greatest confidence that the groups are comparable, ensuring that "like can be compared with like" [2].
- **Elimination of Selection Bias:** It ensures that the investigator has no control over the allocation of participants to either the study or control group, thus completely eliminating what is known as "selection bias" [2].
- **Neutralization of Confounding Factors:** While matching can only account for factors that are known to be important, there may be other factors whose effect is not recognized or cannot be determined [1] [2]. Randomization attempts to distribute both known and unknown confounding factors equally between the two groups [2].

## BLINDING

**Concept:** While randomization balances the groups, it cannot guard against biases that arise from the human element during the evaluation of the trial's outcome [4] [5]. To reduce these problems, a technique known as "blinding" is adopted, which ensures that the outcome is assessed objectively [5]. Note that when an outcome such as death is being measured, blinding is not as essential [6].

**Importance (Biases Prevented):** Blinding prevents systematic errors from three main sources:

1. **Subject Variation:** Participants may subjectively feel better or report improvement if they know they are receiving a new form of treatment [4].
2. **Observer Bias:** The investigator measuring the outcome may be influenced if they know beforehand the particular procedure or therapy to which the patient has been subjected [4].
3. **Evaluation Bias:** The investigator may subconsciously give a favourable report of the outcome of the trial if they are aware of the group allocations [4] [5].

## Types of Blinding:

- **Single Blind Trial:** The trial is planned so that the participant is not aware whether he belongs to the study group or the control group [5].
- **Double Blind Trial:** The trial is planned so that neither the doctor nor the participant is aware of the group allocation and the treatment received [5]. This is the most frequently used method when a blind trial is conducted [6].
- **Triple Blind Trial:** This goes one step further; the participant, the investigator, and the person analyzing the data are all "blind" to the group allocation [5] [6]. Ideally, triple blinding should be used, but double blinding is the most common in practice [6].

**TWO Ethical Considerations in RCTs**

**1. Informed Consent and Risk-Benefit Assessment** Before launching human experiments, it is mandatory that the potential benefits of the experiment are carefully weighed against the risks involved [1]. Participants must give "informed consent," meaning they agree to participate only after being fully informed about the purpose, procedures, and possible dangers or consequences of the trial [1] [2]. Researchers have a strict ethical responsibility to safeguard the rights of the subjects and must observe the ethical principles of beneficence, avoidance of harm, and justice [3].

**2. Withholding Treatment and Placebo Use** A major ethical dilemma arises regarding whether it is acceptable to withhold a potentially beneficial treatment from the control group. It is ethically unacceptable to withhold a treatment whose benefit is self-evident and potentially life-saving (e.g., withholding a blood transfusion for excessive haemorrhage) simply to prove its therapeutic value [4]. While the effects of a new drug are often compared to a concurrent experience using either a placebo or a currently utilized therapy [5], investigators face a profound ethical obligation to intervene, and potentially reduce or eliminate a suspected risk factor, as soon as sufficient evidence of its harm or the new treatment's efficacy accumulates [6].

## Why Experimental Designs Face More Ethical Issues than Observational Studies

- **Observational Studies (Minimal Risk):** In observational designs (such as case-control or cohort studies), nature is allowed to take its own course; the epidemiologist merely observes and measures disease occurrence or past exposures without actively intervening [7]. Because there is no active manipulation or experimental treatment, there is no direct physical risk to the subjects, making ethical problems minimal [8].
- **Experimental Studies (High Risk):** In contrast, an experimental study (like an RCT) involves an active attempt by the investigator to intervene and change a disease determinant or the progress of a disease [7]. Because the investigator deliberately applies, withdraws, or reduces a suspected causal factor (such as a drug or vaccine) in human subjects, the actual health and well-being of the people in the study group may be at stake [9] [10]. This active manipulation necessitates severe ethical constraints that observational studies naturally avoid [9].

## 1. Limitations of Case-Control (Observational) Studies

- **Establish Association, Not Causation:** Case-control studies are analytical (observational) studies designed only to determine whether a statistical association exists between a disease and a suspected factor, and to measure the strength of that association [1].
- **Inability to Distinguish Causes:** A major listed disadvantage of the case-control design is that it does not distinguish between true causes and merely associated factors [2].
- **Statistical Significance $\neq$ Causation:** Finding a statistical association (such as a significant P-value) or estimating risk (such as an Odds Ratio) confirms an association but does not imply causation [3] [4].

## 2. The Role of Experimental Studies (RCTs)

- **Scientific Proof:** To provide "scientific proof" of aetiological (causal) factors, experimental studies are required [5]. The Randomized Controlled Trial (RCT) is the most rigorous epidemiological experiment to evaluate treatments and preventive measures [6].
- **Active Manipulation:** In an RCT, the investigator actively intervenes by deliberately applying, withdrawing, or reducing a suspected causal factor (the independent variable) in an experimental group, and then compares the outcome (the dependent variable) with a control group [7] [8]. This active manipulation allows for the direct observation of cause and effect, which observational studies cannot achieve because they only observe the natural course of events without intervention [5].

## 3. Judging Causality from Observational Data

- Because true experimental trials (RCTs) are often ethically impossible for testing harmful exposures (e.g., making human subjects smoke to prove it causes lung cancer), epidemiologists must rely heavily on inferences drawn from observational studies [9].
- To proceed from demonstrating a mere statistical association to concluding a cause-and-effect relationship, epidemiologists apply additional criteria for judging causality: temporal association, strength of association, specificity, consistency, biological plausibility, and coherence [10] [...]. All these criteria must be utilized together to evaluate the probability of a causal relationship [12].

## BIAS

**Definition:** Bias is defined as any systematic error in an epidemiological study that occurs during data collection, compilation, analysis, and interpretation, which alters the accurate determination of the association between the exposure and the disease [1] [2].

**Types of Bias:** Biases are predominantly classified into three main types based on their source:

### 1. Subject Bias (Error introduced by the study subject) [2]

- **Memory or Recall Bias:** When individuals are asked questions about their past history, "cases" (those with the disease) are more likely to recall the existence of certain events or habits compared to "controls" (healthy persons) [3].
    - *Example:* Patients who have had a myocardial infarction might be much more likely to remember and report past habits or events than healthy individuals [3].
- **Hawthorne Effect (Attention bias):** Study subjects may systematically alter their normal behaviour when they know they are being observed [2].

### 2. Investigator/Observer Bias (Error introduced by the investigator) [2]

- **Selection Bias:** This occurs when the selected cases and controls are not truly representative of the general population, leading to systematic differences in characteristics between the two groups [3].
- **Berkesonian Bias (Admission rate bias):** A specific form of selection bias identified by Dr. Joseph Berkeson. It arises from the different rates of admission to hospitals for people with different diseases, making hospital cases and controls systematically different from each other [2] [...].
- **Interviewer's Bias:** This occurs when the interviewer knows the study's hypothesis and knows who the cases are. This prior knowledge may lead the interviewer to subconsciously question the cases much more thoroughly than the controls regarding the suspected causal factor [4].
- **Neyman Bias (Prevalence-incidence bias):** Bias due to the missing of fatal cases, mild/silent cases, and cases with a short duration of an episode from the study [2].

### 3. Analyzer Bias

- This refers to any systematic error introduced by the analyzer during the evaluation and interpretation of the data [2].

## CONFOUNDING FACTOR

**Definition:** A "confounding factor" (or confounding variable) is a factor that is associated with both the exposure and the outcome (disease), and is distributed unequally between the study and control groups [2] [...]. Importantly, a confounding factor has an independent effect in the causation of the outcome, meaning it acts as an independent risk factor for the disease itself [2] [6].

If not controlled for (e.g., through matching, randomization, restriction, or stratification), confounding can distort or completely falsify the true relationship between the exposure under investigation and the disease [2] [5].

**Suitable Examples:**

- **Smoking in Alcohol-Cancer Studies:** In a study investigating the role of alcohol (exposure) in the aetiology of oesophageal cancer (disease), *smoking* acts as a confounding factor. This is because smoking is highly associated with the consumption of alcohol, and it is also an independent risk factor for developing oesophageal cancer [6].
- **Age in Contraceptive-Cancer Studies:** *Age* could be a confounding variable when investigating the relationship between steroid contraceptives and breast cancer. If the women taking the contraceptives are systematically younger than those in the comparison group, they would naturally be at a lower risk for breast cancer because the disease risk increases with age [7].

## 1. Case-Control Study vs. Cohort Study

- **Direction of Inquiry:** A case-control study proceeds backwards from "effect to cause," starting with the disease and looking back for exposure [1] [2]. A cohort study proceeds forward from "cause to effect," starting with people exposed to a risk factor and following them to see if the disease develops [2] [3].
- **Study Subjects:** Case-control studies involve fewer subjects and are highly suitable for studying rare diseases [1] [2]. Cohort studies require a much larger number of subjects and are inappropriate for investigating rare diseases or rare exposures [2] [3].
- **Outcomes Measured:** Case-control studies yield relatively quick results but cannot calculate exact incidence rates; they only provide an estimate of risk through the Odds Ratio [1] [2]. Cohort studies require a long follow-up period (delayed results) but directly yield incidence rates, exact Relative Risk, and Attributable Risk [2] [3].
- **Cost:** Case-control studies are relatively inexpensive to carry out, whereas cohort studies are highly expensive [1] [...].

## 2. Odds Ratio vs. Relative Risk

- **Odds Ratio (OR / Cross-product ratio):** This is the measure of the strength of association estimated in Case-Control studies [4]. It is calculated as the cross-product ($ad/bc$) because a case-control study lacks an appropriate population-at-risk denominator, making it impossible to calculate true incidence rates directly [4] [5].
- **Relative Risk (RR):** This is the direct measure of the strength of association calculated in Cohort studies [4] [6]. It is the exact ratio of the incidence of the disease among the exposed group to the incidence of the disease among the non-exposed group [6] [7].

## 3. Relative Risk vs. Attributable Risk

- **Relative Risk (RR):** Expressed as a ratio of incidence rates, RR is a direct measure or index of the "strength" of the association between a suspected cause and effect [7] [8]. It is a better index for assessing the aetiological (causal) role of a factor in a disease, but it does not reflect the potential public health impact [9].
- **Attributable Risk (AR):** Expressed as the difference in incidence rates between an exposed group and a non-exposed group (often as a percentage) [6] [10]. It indicates the exact extent to which the disease under study can be attributed to the exposure [11]. AR gives a much better idea of the potential public health impact, showing the amount of disease that could be eliminated if the risk factor is successfully removed or controlled [9] [11].

## 4. Randomization vs. Blinding

- **Randomization:** This is a statistical procedure used at the beginning of an experimental study to allocate participants into "study" and "control" groups [12] [13]. It is considered the "heart" of a randomized controlled trial because it completely eliminates "selection bias" and ensures comparability by distributing both known and unknown confounding factors equally between the two groups [12] [...].
- **Blinding:** Because randomization cannot guard against biases introduced by the human element during the trial, blinding is adopted [15] [16]. It is a technique used during the follow-up and assessment phases to ensure the outcome is evaluated objectively [15] [16]. Blinding prevents systematic errors such as "subject variation", "observer bias", and "evaluation bias", and can be conducted as a single, double, or triple-blind trial [15] [...].