EPIDEMIOLOGY AND PUBLIC HEALTH

Srivani Mandaloju
School of Health Sciences
University of New Haven
300 Boston Post Rd, West Haven, Connecticut, United States-06519
srivani2799@gmail.com

ABSTRACT

Public health is a goal that is carried out via societal action and describes the relationship between health or disease in human populations and other elements associated to health, such as human infections. Epidemiology is a basic science. A large portion of the data needed by public health practitioners to create, administer, and assess successful intervention programs for disease prevention and health promotion has been generated through epidemiology. By recognizing when and how to creatively apply the many epidemiological tactics to address particular health challenges, it is a philosophy and methodology that may be used to understand and address a very broad spectrum of health problems.

Understanding the various study designs and statistical procedures is not enough for an epidemiological investigation; implementation is crucial. To demonstrate and emphasize the fact that understanding study designs and epidemiological methodologies alone is not enough to successfully use epidemiology, the uses and limits of the various epidemiological study designs are discussed. If these designs and approaches are to produce the desired results, they must be used properly, imaginatively, and ingeniously.

The need to make public health judgments in the midst of scientific uncertainty and the possibility that epidemiologic certainty might not give public health action a clear direction. Public health professionals should focus on resolving issues that directly affect public health decisions instead of criticizing epidemiology for failing to address significant public health challenges. Over the past three decades, the subject of epidemiology has grown significantly as researchers have found novel applications for established study designs and methodologies. As more and more inventive epidemiologists create cutting-edge new strategies and procedures, it is projected that the field of epidemiology will broaden even further in the future.

Keywords- epidemiology; study designs; health problems; public health

I. INTRODUCTION

In order to manage health issues, epidemiology is the "study of the distribution and determinants of disease frequency or occurrence of illnesses in human groups" [10]. The history of epidemiology dates back around 400 years, during which time advancement was moderate and uneven. John Snow demonstrated that cholera was spread by faecal contamination of drinking water in the 19th century, whereas James Lind employed experimental studies to determine the etiology of scurvy in the 18th century and John Graunt detailed the pattern of mortality in the 17th century.

A. Objectives of epidemiology

- To pinpoint the underlying causes of diseases and important risk factors that raise a person's risk for developing them.
- To assess the severity and burden of disease in the neighborhood.
- To investigate the course and prognosis of disease.
- To assess both existing and newly created preventive, therapeutic, and delivery methods for medical care.
- To serve as a starting point for the formulation of public policy pertaining to environmental issues, genetic concerns, and other factors pertaining to illness prevention and health promotion [5].

B. Types of epidemiology

Descriptive epidemiology, which describes illness patterns, is a subfield of epidemiology. Its primary goals include generating hypotheses regarding the causes of diseases, monitoring public health, and gauging the effectiveness of intervention initiatives. Its purpose is to find and count disease cases in populations and carry out easy research. Case report, Case series, cross-sectional study, and ecological study are other classifications. Descriptive epidemiology looks for trends by looking at individual variables like occupation, diet, religious beliefs, age and gender, place and time, and occupation. When a disease epidemic arises, these traits are carefully taken into account since they offer crucial hints about the origin of the outbreak. Governments, non-governmental organizations, health maintenance organizations, hospitals, and independent researchers collect data from the US Census, health surveys, death and birth certificates, cancer registries, hospital discharge registries, and infectious disease reporting that is already in place, well-established, and low-cost but that may contain inaccurate information, exclude data we want, cause reporting delays, or involve complicated methodology.

A case report is a description of a patient who has a rare ailment or who is experiencing multiple conditions at once. Case series are comparable to individual cases, however they are collections of numerous related cases. In a cross sectional study, data are gathered on the existence or level of one or more variables of interest, such as exposure or outcome, as they are present in a defined population at a certain point in time. Ecologic research examines associations between an exposure and a result across populations as opposed to in a single subject. A key restriction was that associations at the group level might not hold true at the individual level [1].

Analytical or scientific epidemiology is another form that assesses theories on the origins of disease and the effectiveness of intervention initiatives. It compares groups and works methodically to find any connections between cause and prevention. It is divided into cohort studies, case control studies, experimental studies, and clinical trials. It has been stated that epidemiology cannot ever demonstrate that a specific exposure led to a specific outcome. However, epidemiology frequently offers enough proof to implement suitable control and preventative measures.

A clinical trial is thought of as an investigation of the patterns, causes, and effects of health and disease in people as well as the connection between exposures or treatments and medical results. When determining the type of exposure, the researcher uses data from specific individuals. To assess the impact of the treatment, participants are monitored. In an experimental study, people are

randomly assigned to exposures, and their health over time is monitored to see if they become ill or recover from it. A relationship between an exposure and a health consequence is to be identified and measured. In a case control study, researchers first enroll a group of participants who have the disease and compare them to controls who do not have the disease or condition. In a cohort study, which is comparable to an experimental study, the epidemiologist keeps note of whether each study participant has been exposed to the target disease or not before following up to see whether they do.

II. MEASURES OF DISEASE FREQUENCY

Finding probable causes of disease and figuring out efficient approaches for prevention and care requires measuring the prevalence of disease in a community and figuring out how it might change over time. It has four parts, including population, disease cases, population size, and time.

A. Population:

A group of individuals with a common trait was used as the base group from which disease frequency was calculated and epidemiological usage was used. There are two categories of populations: fixed and dynamic. Fixed populations are those whose membership is based on a particular event and is therefore permanent, such as veterans of the Vietnam War. A dynamic population was characterized as a group that has a cyclical membership, such as Boston residents, graduate students, or religious organizations.

B. Cases of Disease:

This is the numerator of all frequency measurements and is regarded as any adverse health outcome, including an illness, a birth defect, or an injury. Clinical records, diagnostic testing, disease registries, surveillance programs, and self-reports are examples of disease ascertainment techniques. In this, high-quality data is essential.

C. Population size:

It serves as the denominator for all calculations. This is based on a population that has been identified as being important for comparing the disease across populations as the sheer number of cases makes such a comparison impossible.

D. Time:

It is essential for all indicators of the frequency of diseases. A single point over a period of time can be used to evaluate the prevalence of a disease. For instance, a follow-up period from birth to age 10 for children residing in Boston.

Ratio, proportion, and rate are three general categories of measurements. In proportion, numbers must be connected, and the numerator is a subset of the denominator, which is frequently stated as a percentage. Ratio is the division of one number by another, but the numbers need not be related. Rate, which is the most frequently misunderstood quantity, is the division of two numbers by which time is a fundamental component of the numerator.

Prevalence, cumulative incidence, and incidence rate are epidemiological measures. Prevalence indicates the number of disease cases that have actually occurred in a population or the percentage of people who have the disease over a predetermined time period while living in a state where the complete population is included in the denominator. The percentage of the population who are afflicted with the disease at any one time is known as point prevalence. It is not a unit.

$$\text{Prevalence} = \frac{\#\ \text{EXISTING cases at time point}}{\#\ \text{TOTAL POPULATION at time point}} \tag{1}$$

The prevalence of new disease cases in a population during a predetermined time period is measured by incidence. It could entail a change in state, such as going from being healthy to sick, from being living to being dead, or from being sick to being healthy. Population at risk is included in the denominator, but those with the condition being measured and immunological status are excluded. Cumulative Incidence (CI) and Incidence Rate are the two types of incidence (IR). CI measures the number of new cases of disease that occur in a population (the percentage of the population that are at risk for disease) during a given period of time. No follow-up will be lost during the whole set time period for any of the population's members. It is not a unit.

$$\text{Cumulative Incidence} = \frac{\#\ \text{NEW cases during time period}}{\#\ \text{POPULATION AT RISK at start of time period}} \tag{2}$$

Attack rate and case fatality rate are particular subtypes of cumulative incidence. Attack rate is the percentage of people who are exposed to an infectious agent over time and develop an infection. The percentage of patients who pass away from their illness is known as the case fatality rate.

$$\text{Attack Rate} = \frac{\text{New cases of disease during specified time period}}{\text{Population at risk at beginning of time period}} \tag{3}$$

$$\text{Case Fatality Rate} = \frac{\#\ \text{Deaths due to disease X during specified time period}}{\#\ \text{People with disease X}} \tag{4}$$

Limitation of CI presupposes that it is not a perfect indicator in a changing population or in a population that remains constant but loses members over time. Time of occurrence is not taken into account.

The term "incidence rate" refers to the rate at which new instances of disease appear in a population; it does not imply that the population being studied has been under observation for the whole time period. The denominator inherently includes time. "Person-time" is the denominator, and the units are "time" or "cases/person-time."

$$\text{Incidence Rate} = \frac{\#\ \text{NEW cases during time period}}{\#\ \text{Total person}-\text{time of observation in population at risk}} \tag{5}$$

E. Person time:

Duration of each at-risk person's surveillance. It only builds up among those who are at risk while they are being monitored, and it stops when the individual gets the condition being studied, passes away, is lost to follow-up, or doesn't match the requirements for eligibility. Different quantities of person-time are contributed to the denominator by each person. P and IR relationship was described as

$$P = IR * D \tag{6}$$

Where D=average duration of disease; time from diagnosis to recovery or death.

As opposed to IR, which does not imply comprehensive follow-up and can take into consideration when a disease has developed, CI has the advantages of being easily calculated, understood, and applied to assess individual risk. The disadvantages of CI and IR include that CI assumes complete follow-up for every individual and ignores the time of disease onset while IR requires non-intuitive interpretation, cannot be quantified, and can be challenging to determine person-time. IR is employed in dynamic or fixed populations, with short or no lost to follow-up intervals in fixed populations [9].

III. COMPARING DISEASE FREQUENCIES

The comparison of prevalence, cumulative incidence, and incidence rates according to exposure status is at the heart of epidemiology. It is compared because there is a link between exposure and disease if a disease affects one group more frequently or less frequently than the other. Calculating the difference between two measures of disease frequency is an absolute measure, while calculating the ratio of two measures of disease frequency is a relative measure, respectively.

A. Absolute measures of Association

Prevalence difference is the difference in prevalence between index and comparison groups.

$$PD = P_E - P_U \quad (7)$$

Risk difference is the difference in cumulative incidence between index and comparison groups.

$$RD = CI_E - CI_U \quad (8)$$

Rate difference is the difference in incidence rate between index and comparison groups. Units are person-time.

$$RD = IR_E - IR_U \quad (9)$$

If there is no association between the exposure and disease then RD = 0, if the exposure is associated with increased risk of disease then the RD > 0 and if the exposure is associated with decreased risk of disease then RD < 0.
Prevalence and Cumulative incidence data can be calculated by using 2x2 table.

Table 1: 2x2 table for the calculating Prevalence and Cumulative Incidence

	Disease	No Disease	Total
Exposed	a	b	a+b
Unexposed	c	d	c+d
Total	a+b	b+d	a+b+c+d

Example 1:

Among 13,422 women with hypertension, 117 had an MI over 10 years of follow-up. Among 106,541 women without hypertension, 125 had an MI during the same follow-up period.

$$RD = CI_E - CI_U$$

Risk difference = CI in exposed women – CI in unexposed women
= (117/13,422) – (125/106,541)
= 0.0087 – 0.0012
= 0.0075
= 75/10,000 over 10 years of follow up

Example 2:

Among 13,422 women with COPD, 117 has an MI over 15 years. From them 106,541 women without COPD, 125 has an MI over. From this data exposure and disease can be identified by using 2x2 table.

Table 2: 2x2 table for calculating Cumulative Incidence

	Disease (COPD)	No Disease (No COPD)	Total
Exposed (MI)	117	125	242
Unexposed (No MI)	13,305	106,416	119,721
Total	13,422	106,541	119,963

Calculation:
Cumulative Incidence
For COPD = 117/13,422
For No COPD = 125/106,541
Risk Difference
For COPD = (117/13,422) – (125/106,541)
= 75/10,000 over 15 years

From this it can be interpreted that, women with hypertension had 75 more MIs per 10,000 women over a 10-year period compared to women without hypertension.

B. Relative measures of Association

Prevalence Ratio is the ratio of prevalence between index and comparison groups.

$$PR = P_E/P_U \quad (10)$$

Risk ratio is the ratio of cumulative incidence between index and comparison groups.

$$RR = CI_E/CI_U \quad (11)$$

Rate ratio is the ratio of incidence rate between index and comparison groups.

$$IRR = IR_E/IR_U \quad (12)$$

Relative risk is used to represent all classes of relative measures of association often used interchangeably with risk ratio or cumulative incidence ratio. If there is no association between exposure and disease, then RR = 1, if it is associated with increased risk of disease, then RR > 1 and if associated with decreased risk of disease, RR < 1.

Example 3:

Among 13,422 women with COPD, 117 has an MI over 20 years of follow-up. Among 106,541 women without hypertension, 125 has an MI during the same follow-up period.

$$RR = CI_E / CI_U$$

Risk Ratio = CI in exposed women / CI in unexposed women
= (117/13,422) / (125/106,541)
= 0.00872 / 0.00117
= 7.45

For example, among 13,422 women with hypertension, 117 had an MI over 10 yrs. Among 106,541 women without hypertension, 125 had an MI over 10 yrs.

Table 3: 2x2 table for calculating Cumulative Incidence

	Disease (Hypertension)	No Disease (No Hypertension)	Total
Exposed (MI)	117	125	242
Unexposed (No MI)	13,305	106,416	119,721
Total	13,422	106,541	119,963

Calculation:
Cumulative incidence
For Hypertension = 117/13,422
For No Hypertension = 125/106,541
Risk Difference = (117/13,422) – (125/106,541)
= 75/10,000
Risk Ratio = (117/13,422) / (125/106,541)
= 7.45

From this, it can be interpreted that women with hypertension were 7.45 times as likely to have an MI over a 10 year period compared with women without hypertension.

Excess Relative Risk = (RR-1) x 100%
If RR = 7.45, women with hypertension had 645% increased risk of having an MI compared with women without hypertension.
Excess RR is said when,
RR = 1.0, Excess RR = 0%
RR = 2.0, Excess RR = 100%
RR = 1.6, Excess RR = 60%
RR = 0.5, Excess RR = -50% or 50% reduced risk

When determining the amount of sickness that can be directly linked to an exposure, we consider the population attributable proportion. An association is a measurable connection between exposure and an illness that results from comparing risks or rates and suggests that exposure might contribute to a disease. We refer to exposures as risk factors or preventative factors when they are connected to variations in disease risk. Observed correlations are not always indicative of a causal connection.

Epidemiological research cannot pinpoint the origin of an illness in a specific person. Instead, they establish the connection between a specific exposure and the prevalence of disease in communities. Based on the known link between an exposure and a disease, among other things, epidemiologists infer causation. A determination is causal interference.

Assuming a causal association between exposure and disease, the proportion of disease in the total population that would have been prevented if the exposure had not occurred.

$$\text{Difference measures} = \frac{R_T - R_U}{R_T} \text{ x } 100 \quad (13)$$

Where R= P, CI or IR

$$\text{Ratio measures} = \frac{[P_e(RR-1)]}{[P_e(RR-1)+1]} \text{ x } 100 \quad (14)$$

Where RR = PR, RR or IR and P_e = Proportion exposed

Example 4:

Using RR, for calculating proportion of MI cases in the total study population that would have been avoided if the women did not have hypertension we may use population attributable proportion.

Table 4: 2x2 table for calculating proportion exposed

	Disease (Hypertension)	No Disease (No Hypertension)	Total
Exposed (MI)	117	125	242
Unexposed (No MI)	13,305	106,416	119,721
Total	13,422	106,541	119,963

P_e = Proportion exposed
= Total population Exposed/Total population

= 13,422 / 119,963
= 0.112

$$\text{Population attributable proportion} = \frac{[P_e(RR - 1)]}{[P_e(RR - 1) + 1]} \text{x } 100$$

$$= \frac{[0.112\ (7.45\ -\ 1)]}{[0.112\ (7.45\ -1\ ) + 1]}\ \text{x } 100$$

= 41.9 %

It can be interpreted that, if the association between hypertension and MI is causal, 41.9% of MIs among women in the Nurses' Health Study would have been avoided if they had normal blood pressure (instead of hypertension) during the 10 year study period.

IV. SOURCES OF PUBLIC HEALTH DATA

There is a plethora of knowledge available in the United States about the health of populations all around the world. The majority of the information is gathered routinely or through specialized surveys by governmental and non-governmental organizations. A wide range of illnesses and ailments, including acute illnesses and injuries, chronic illnesses and impairments, birth abnormalities and other unfavorable pregnancy outcomes, are all subject to information availability. There is also information available on factors that affect a person's risk of illness, such as dietary practices, immunizations, and use of tobacco, alcohol, and drugs, as well as the effects of these illnesses on the use of health services, such as hospitalizations, visits to doctors' offices, and use of emergency and outpatient hospital departments. There are numerous sources of global data that include details on births, deaths, and important health metrics.

The U.S. population census, which counts the entire population once every ten years, as well as data on demographic traits and risk and rate denominators, are important sources. In the US, vital statistics on births, deaths, marriages, divorces, and fetal deaths are all seen as being extremely essential. You can get general information about birth weight, gestational age, death reason, and underlying disorders.

The National Survey of Family Growth collects data from all 50 states on men and women's marriage, divorce, family planning, and infertility. The National Health Interview Survey collects information on key health issues, including as acute illnesses and injuries, chronic illnesses and disabilities, and the use of medical services, at the national level. The US population's health and diet are studied as part of the National Health and Nutrition Examination Survey. That include interviews and medical examinations. The behavioral Risk Factor Surveillance System conducts telephone surveys on health risk behaviors related to chronic disease, injuries, and death. These behaviors include using screening and preventive services, smoking, drinking alcohol, engaging in physical activity, eating fruits and vegetables, using seatbelts, and controlling one's weight. The National Health Care Survey is a collection of surveys about the usage and effectiveness of healthcare, as well as the effects of medical technology, in a range of settings, including emergency rooms, hospices, home health agencies, and doctor's offices. Weekly statistics on the prevalence of more than 60 notifiable diseases, such as AIDS, HIV infection, botulism, gonorrhea, hepatitis, syphilis, plague, and malaria, are gathered by the Notifiable Diseases Surveillance System. The CDC posts findings in Morbidity and Mortality: The Weekly Report (MMWR).

The National Immunization Survey gathers data on US children's immunization status. Diphtheria, tetanus, pertussis, polio, measles, mumps, rubella, hepatitis B, and influenza are among the vaccinations. Information on mental health, alcohol consumption, and illicit drug use is gathered by the National Survey on Drug Use and Health. In 18 population-based registries across the US, the Surveillance, Epidemiology and End Results Program gathers data on cancer prevention, diagnosis, and treatment. Birth Defects Surveillance and Research Programs gather information, carry out research, and inform the public on birth defect prevention.

The World Health Statistics Annual compiles international morbidity and mortality data on 194 WHO member states, and the WHO International Agency for Research on Cancer (IARC) collects data on cancer incidence and mortality from many countries around the world where the incidence is highest. The demographic yearbook published by the United Nations collects data on 230 countries and areas of the world on population size, distribution, and growth, births, deaths, marriages, and divorces. For the purpose of understanding data, it is critical to understand the precise population that the data collection system covers. Understanding the time period that the data collection system covers and how frequently data are updated are the two main factors to take into account, but the data that is currently available is one to two years behind current events, and every data collection system contains some incomplete and inaccurate data [3].

V. EPIDEMIOLOGICAL STUDIES

Studies might be either descriptive or analytical. Incidence, prevalence, and experience are provided by descriptive studies, which make no attempt to quantify relationships. It is based on surveys, in-depth case studies, case reports, and occurrence. A case report is a comprehensive account of one patient who typically has a novel or uncommon symptom or issue. A case series is a thorough account of a number of patients who have the same symptom or issue; typically new or unique. The number of new cases of a disease during a certain period in a specific population is described by incidence studies. These are helpful for identifying and describing novel diseases, new symptoms of existing diseases, identifying adverse drug reactions, and giving insight into the pathophysiology of diseases.

Analytical studies quantify the connection between an intervention's effect and its exposure or therapy. According to ethical standards, experimental studies are those in which the researcher assigns exposures; randomized controlled trials are those in which the allocation is random, while non-randomized controlled studies are those in which it is not. If the researcher does not assign exposures, the study is an observational one; if a comparison group is present, the study is an analytical one; otherwise, it is a descriptive one. They are divided into four categories based on the analytical study method: cohort, case-control, cross-sectional study, and ecological.

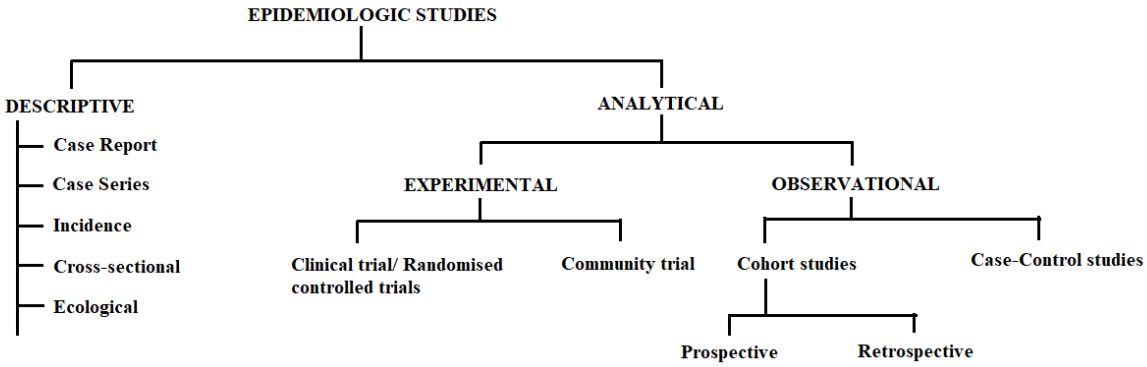

Figure 1: Classification of Epidemiological studies

A. Descriptive epidemiology

i. Case Report

A case report is an in-depth account of a disease's manifestation in a single individual. Unusual aspects of the case may raise a fresh theory regarding the origins or pathophysiology of the illness. It frequently discusses exceptional examples that cannot be explained, which demonstrate disease variety, unexpected occurrences, and situations in which a single patient has two or more unexpected disorders. Case reports are regarded as the weakest form of evidence, but they also serve as the initial source of information because this is where novel problems and concepts are raised. For instance, Infant Acquired Immunodeficiency and Potential Blood Product Transmission.

These reports are helpful in detecting new pharmacological side effects and prospective applications that may be harmful or helpful can help identify new trends or diseases, and it can also identify rare disease presentations. These cases may not be generalizable; they may not be based on systematic investigations; causes or relationships may be explained by other factors; and they may be viewed as highlighting the weird or concentrating on false information.

ii. Case Series

A case series is a description of the traits of a number of individuals who have the same illness or condition. It is a group of subjects that have a common trait and is used to define certain clinical, pathophysiological, or operational features of a disease, its treatment, exposure, or diagnostic process. A case series, as opposed to a case report, is an abstraction of a number of examples gathered over time that led to theories about potential new diagnoses, treatments, or side effects that could be rigorously evaluated in the future. Case series can be structured formally or informally, consecutively or not, retrospectively or prospectively, based on a clinic or a community, and based on exposure or outcome.

From this, can identify novelties that are uncommon, generate new hypotheses, and cost less than formal investigations. There is no interference with the choice of treatment, little time is required for research execution, and it may be helpful when a randomized controlled trial is not appropriate or feasible. Limitations include no comparison group, frequently poor data collection, difficulty examining disease etiology, difficulty accessing disease frequency, and lack of external validity due to the possibility that cases are not representative.

iii. Incidence studies

The ideal incidence study measures all population members' exposures, confounding factors, and outcome times. The study may be referred to as a "cohort study" or a "follow-up study" when the source population has been formally identified and counted; the former terminology will be used in this context. Studies where the source population has been identified but a cohort has not been explicitly counted by the researcher, such as "descriptive" studies of national death rates, also fall under the category of incidence studies. Furthermore, there is no real difference between incidence studies using a large population and incidence studies using exposure-based sampling.

Incidence studies frequently employ three disease occurrence metrics. A second measurement is the "occurrence proportion" or average risk, which is the percentage of study participants who encounter the outcome of interest at any point throughout the follow-up period. The person-time "incidence rate" is likely the most prevalent measure. The ratio of participants who experienced the outcome to those who do not is known as the "incidence odds," and it can be used as a third possible measurement. The number of incident disease cases serves as the numerator for each of these three indicators of disease occurrence. Whether their denominators are survivors, persons at risk, or person-time at risk makes a difference.

The three ratio measures of effect employed in incidence studies are the "rate ratio," "risk ratio," and "odds ratio," which correspond to these three measurements of disease occurrence.

iv. Cross-sectional

In this, a sample of people from a previously identified group are chosen and contacted at a specific time to gather data on the relevant exposures and outcomes simultaneously. They are typically carried out to estimate the prevalence of the desired result for a certain population, sometimes for public health planning purposes. Cross-sectional government surveys are common.

It is relatively quick, cheap, extremely generalizable if centered on general population, can be avoided difficulty with sequential inferences for unchangeable, long-term, and historic exposures. Studies that are cross-sectional have issues when exposure is a variable attribute. These investigations take into account unchangeable traits, measurements of long-term exposure, and historical exposure. Problem with employing frequent occurrences because prevalence mixes incidence and duration, it is not optimal for etiologic study. Instances with a long duration are more likely to be discovered as prevalent cases in cross-sectional surveys.

v. Ecological

A population-level component is examined in ecological studies in relation to disease rates. Instead of the individual, the group is the unit of observation. There is no actual connection between exposure and disease because exposure is calculated as the average for

a community, not an individual. Our focus, however, is on consequences at the individual level. The group-level association may not translate to the individual level, according to the ecologic fallacy.

Ecological studies are typically unable to control for confounding factors, or other variables that would explain the observed link. Could there be another factor contributing to the link between cigarette sales and heart disease, for instance? What elements may function as confounders? Relationships with complexity can be hidden.

It uses readily available data, is quick and affordable to undertake, is effective for advancing early understanding, covers a greater variety of exposures than other types of studies, and may be used to investigate ecological linkages. Consider the impact of motorcycle helmet rules on the state's motorcycle fatality rate. Using a correlation coefficient or linear regression makes analysis simple [4].

B. Analytical Epidemiology

i. Experimental studies

Clinical trials and community trials are the two main categories of randomized trials, with randomized clinical trials being significantly more prevalent.

a. Randomized controlled trials or clinical trials

Randomized trials are epidemiological studies that compare two or more treatment groups directly, with one of the groups acting as a control for the others. The study participants are divided into several treatment groups at random, and the effects of the various therapies are tracked in all groups over time. Randomized trials offer the clearest proof of causality. A randomized clinical trial is a study in which participants are actual people. The objective is to examine an intervention to stop the progression of a disease or to find an effective treatment for an illness. Randomized clinical trials are frequently used to compare the effectiveness of new drugs to placebos or to standard treatments, but they are also used to assess the efficacy of other therapeutic interventions, such as a novel surgical technique, a dietary plan, or an exercise program for people with pre-existing disease.

b. Community trial

Although a community trial is also an experiment, it varies from clinical trials in that it uses the entire community as the unit of observation rather than a single patient. For instance, water fluoridation was assessed by experimentally deciding whether to fluoridate or not fluoridate the public water supply for entire communities.

The impact of media programs that promote increased exercise, reduced tobacco use, and other lifestyle changes to prevent heart disease has been examined in a number of community trials. Only when equipoise exists is an experimental study ethically and appropriately conducted. When researchers are in an equilibrium condition, they genuinely do not know whether treatment is superior to another yet genuinely feel that depriving study participants of treatment will not hurt them.

ii. Observational studies

a. Cohort studies

A cohort study is a sort of long-term research that enlists a group of people who have similar traits over time. People with the same occupation, belonging to the same race or demographic group, or even those afflicted with the same illness, might all be included. Observational studies is of two types, prospective and retrospective.

A prospective cohort study is a sort of cohort study in which the researchers plan the study, recruit participants, and gather background information on each participant before beginning to produce notable outcomes. A prospective cohort study includes the Framingham Heart Study. One significant benefit of a prospective cohort study is that it spares researchers from dealing with the moral dilemmas raised by randomized control trials i.e. who receives a placebo and who gets the actual treatment. Multiple diseases and outcomes can be researched at once. It is simple to calculate the incidence and prevalence of a condition. Negative aspects include the possibility of selection bias and confounding variables. Cohort studies can be time- and resource-intensive. Large sample sizes are frequently needed.

Retrospective cohort study compares the incidence of disease among patients by grouping them according to their exposure status. In this instance, though, researchers go back in time to find a group that wasn't initially exposed and analyze the frequency of their exposure. Retrospective cohort studies, such as Lane-1926 Claypon's investigation of breast cancer risk factors, were the first to be acknowledged. An ambi-directional cohort study is referred to as being both prospective and retrospective. This indicates that the study contains both prospective and retrospective aspects.

b. Case-Control studies

An investigation into the connection between a risk factor and a disease that contrasts individuals with the disease or result of interest (cases) with patients without the condition (controls) and analyzes exposure to risk factors across time in each group.

Case control studies are observational in nature because no interventions or attempts to change the course of the disease are made. The objective is to retroactively ascertain how much each of the two groups of people—cases and controls—were exposed to the relevant risk factor. These investigations aim to calculate odds.

Because the ailment or disease has already manifested, conducting the study takes less time and is more effective for investigating rare conditions or diseases than cohort studies. Look at several risk factors at once. Initial investigations that establish a link can be helpful and can provide answers to questions that other study designs were unable to provide.

Because retrospective studies rely on memory and people with certain conditions will be more motivated to recall bias, there are more issues with data quality in these types of studies.

Finding a proper control group can be challenging, and absolute measures of association cannot be calculated. These drawbacks make them unsuitable for evaluating diagnostic tests because it is obviously obvious that the cases have the condition and the controls do not [11].

C. Components of a study

i. **Population:** The source population is the one that is worth knowing. The study population is made up of the participants in the study's source population.

ii. **Exposure:** A factor of interest, such as a constitutional, environmental, or behavioral factor, upon which an outcome depends. For instance, female sex, genetic polymorphism, and small stature.

iii. **Outcome:** A description of the phenomenon under research, which may be a disorder, flaw, harm, occurrence, or state. Error minimization is the main objective. For instance, a woman with breast cancer who was discovered in a cancer registry.

iv. **Potential Confounders**: Unrelated factors that may be behavioral, environmental, or constitutional in nature and influence an outcome. Confounders may cause the true relationship between exposure and disease to be misrepresented. Confounders must be located and managed in some way.

v. **Analysis:** Review of study results, calculation of disease frequency and association measurements.

vi. **Communication of Findings:** Share your findings with the right people in your community, even if there is no correlation.

VI. EPIDEMIOLOGICAL PROBLEM ORIENTED APPROACH

The epidemiologic problem-oriented approach (EPOA) methodology aids in the creation of systematic and structured knowledge bases, which are essential for the creation of models in the field of disease epidemiology. The knowledge base for numerous diseases was developed using the EPOA methodology. It is made up of two triads with six pillars each, linked by the diagnostic technique, such as the Problem identification or characterization and the Problem management or solution or mitigation triads.

The triads are divided into their corresponding pillar variables and parameters using data from numerous sources. The causative agent and its features are named in the agent pillar. The host pillar identifies and describes the host, whereas the environment pillar describes the host's and the agent's physical, biological, and socioeconomic settings. The therapeutics or treatment pillar takes the available therapies for the specific ailment into account. The prevention or control pillar takes these actions into account. In contrast, the health promotion or health maintenance pillar takes into account population health maintenance strategies.

Conceptual, *in vivo* or *in vitro*, systems analysis, mathematical, or computational models of the illness are only a few examples. The knowledge base created using the EPOA methodology offers a well-structured and organized source of data that is used in the estimation of variables and parameters as well as analysis including biological, mathematical, statistical, and computer simulations, all of which are essential in epidemiologic modeling of disease. The creation of models that can aid in public health decision-making has increasingly relied on the EPOA [8].

This is an illustration of a method focused on epidemiological problems in tuberculosis. In this model, the host is a person, the agent is Mycobacterium tuberculosis, and the environment is made up of three categories: biological (nutrition and gender) and physical (temperature and air pollution) and socioeconomic (unemployment and education) aspects.

The clinical signs, symptoms, and tests that are utilized to identify the condition are addressed by a diagnostic link between two triads. The second triad covers problem classification, treatment, and prevention or control. Treatment contains information on quadruple therapy, vaccine, ventilation, and diagnosis. Treatment adherence and outcome are addressed in health maintenance knowledge awareness.

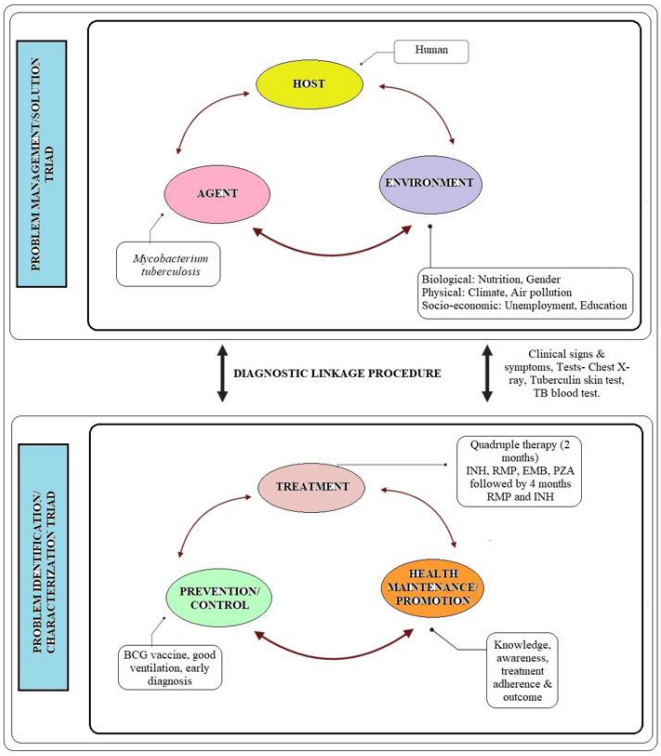

Figure 2: Example of Epidemiological problem oriented approach for Tuberculosis

VII. MINIMIZING BIAS

A. Blinding

A technique called blinding or masking prevents study participants and researchers from knowing whether a participant has been assigned to the treatment or control group.

i. **Single-blind:** Participants in the study are unaware of whether they are receiving treatment or not.

ii. **Double-blind trial:** neither the participant nor the study investigator is aware of who is receiving treatment or none at all.

iii. **Triple-blind:** No one involved in the study, including the participant, the researcher providing the treatment, and the researcher evaluating its results, is aware of whether a subject is receiving the treatment or not.

It's not always possible for blinding in specific treatments, including surgery and therapy, diet and exercise regimens, and major side effects.

B. Placebos

Because it makes the experiences of the exposed and unexposed groups as comparable as feasible, the placebo, one way of blinding, is used. Surgery, exercise, diet, and other therapies cannot be done with a placebo. They frequently employ the accepted standard of care rather than a placebo and are not always ethical. The placebo effect is a positive outcome brought on by a placebo drug or treatment that cannot be traced to the placebo's inherent qualities and must instead be the result of the patient's belief in that particular course of treatment.

C. Compliance

Throughout the experiment, it adheres closely to the requirements of the study protocol. Protocol changes may be necessary due to adverse effects, illness, level of interest, or duration of follow-up. For instance, in a research on physicians' health, compliance meant taking a medication daily.

Noncompliance tends to homogenize the comparison and treatment groups, which makes it harder to distinguish between them. It should be kept in mind that there will be a comparison with the variations between the treatment and control groups. Even if there is a difference, it won't be visible if started to resemble one another more. In order to achieve good compliance, various steps need to be taken. Non-compliance has a bias in favor of "null."

VIII. SOURCES OF ERROR IN EPIDEMIOLOGICAL RESEARCH

Systematic and Random errors fall into two categories. Systematic error has a known source, whereas random error is caused by chance. Both random and systematic mistake are possible. The systematic flaw in the study's design or execution known as bias results in an erroneous estimate of the association.

Sources can be brought on by the researcher or study participants during study planning or execution. It can happen in ecological, experimental, cohort, case-control, and cross-sectional studies. Not a defining trait of the study population. Very few studies are free from bias or mistakes.

Selection bias and information bias cannot be changed in the analysis, while confounding can be fixed to a certain extent. They can create the impression of a link where there is none or hide an association that actually exists. Limitations in study design, restrictions on study conduct during data collecting, and a critical evaluation of the study's results after it has been finished are possible remedies. It is important to talk about the types, origins, direction, magnitude, impact on the findings of the investigation, and study interpretation.

The null value is biased in the direction of associations that are positive and preventative. It is undervalued when true association occurs. Positive and preventative associations both have a bias against the null value. The strength of true association is inflated.

A. Types of Biases

i. Selection Bias

It is a bias that might happen if study participants are not representative of the population being studied. Selection bias relates to who is included in the study. Results of methods used to choose study participants at the time of recruitment and during the process of keeping study participants. Results in an observed correlation that is different from what the source population that was the focus of the investigation would have revealed.

It happens in cohort, experimental, and case-control studies. The choice/participation of cases and controls in case-control studies is correlated with exposure status. Selection/participation of exposed and unexposed patients in cohort/experimental studies is connected to disease status. Participation varies depending on exposure and illness. Due to the fact that exposures and outcomes have already taken place by the time a subject is chosen for a research, it is more likely to happen in case-control or retrospective cohort studies. Selection bias is divided into three categories: cohort experimental on differential loss to follow-up, case-control cohort on differential participation, and cohort case-control cohort on differential control selection.

When controls are more or less likely to be chosen depending on whether they are exposed or not, control-selection bias can develop. Because controls do not precisely represent the same source population as the cases, they will not adequately reflect the exposure distribution in the source population from which the cases originated. If one's motivation or capacity to engage is influenced by both exposure and disease condition, there may be differential participation bias. For all categories, it achieves strong participation rates. If study participants leave the study for reasons linked to both exposure and disease, there is a differential loss to follow-up bias. Since results cannot be predicted without thorough follow-up, participation rates must remain high. By correcting selection bias after it has already happened, bias can be decreased. By conducting and designing studies carefully, selection bias must be avoided. It is uncontrollable during the analysis.

ii. Information Bias

If the data gathered from or about study participants is inaccurate, bias may result. The information that enters your study has a role in information bias. Occurs after study participants have enrolled. Study participants may be wrongly categorized as exposed or unexposed, or as ill or not ill, as a result of variations in how information is acquired. Results in a connection that is different from what would have been discovered if all study participants had been accurately categorized.

Both cohort studies and case-control studies have it. In cohort studies, exposed and unexposed groups are employed, but in case-control studies, different methodologies are used to collect information from cases and controls. When exposure information on exposure and disease is different, bias will always result. Because exposures (case-control studies) and outcomes (retrospective cohort studies) have already happened by the time a subject is chosen for a study, they are more likely to take place in these types of studies.

There are three types of information bias. Measurement error, interviewer bias, and recall bias. Recall bias can happen when those who have an illness report their exposure in a different way than those who do not. In a retrospective cohort research, exposed participants are more or less likely to remember prior diseases than unexposed participants, and cases are more or less likely to recall prior exposures than controls in case-control studies. Utilizing controls who are ill, can help encourage comparable recollection and reduce recall bias. To encourage uniformity and specificity, use standardized, closed-ended surveys. To determine exposure, examine

previous data or make use of biological measurements. If there is a systematic variation in how information is gathered, recorded, or interpreted, interviewer bias may result. In cohort and experimental studies, the treatment or exposure status of the participant has an impact on the interviewer, much as it does in case-control studies where participants are either cases or controls. Utilizing blinding to prevent interviewers from studying hypotheses or knowing whether someone is exposed is one way to address interviewer bias. Use high-quality, structured, closed-ended surveys to encourage precision and uniformity and the use of suitable response categories. Examine the data that has already been collected, and give interviewers proper and thorough training.

If study participants are assigned to the incorrect exposure or disease category, measurement error may result. This sort of bias occurs the most frequently across all study types. Sources include self-reports, mistakes on death certificates, medical records, and other documents, data entry mistakes, and definitions of diseases or exposure that are too general. Effects include differential and non-differential biases toward or away from the null. Utilizing blinding to prevent interviewers from studying hypotheses or knowing whether someone is exposed is one way to address interviewer bias.

Use high-quality, structured, closed-ended surveys to encourage precision and uniformity and the use of suitable response categories. Examine the data that has already been collected, and give interviewers proper and thorough training. If study participants are assigned to the incorrect exposure or disease category, measurement error may result. This sort of bias occurs the most frequently across all study types. Sources include self-reports, mistakes on death certificates, medical records, and other documents, data entry mistakes, and definitions of diseases or exposure that are too general. Effects include differential and non-differential biases toward or away from the null. Information bias must be avoided by thorough study design and conduct because there is nothing that can be done to correct information bias after it has already occurred. It is uncontrollable during the analysis [6].

iii. Confounding

Populations' unequally distributed traits are to blame. A confounder is an unrelated variable that influences the variables under study in such a way that the results do not accurately reflect the relationship between the variables. All variables that are not the independent variable but may have an impact on the experiment's findings are referred to as extraneous variables. The independent variable is the cause. Its value is unaffected by the other study variables. Effect is the dependent variable. Changes in the independent variable affect its value.

It is based on the assumption that there are systematically different groups being compared, which falsifies the actual link between an exposure and a disease. When the exposed and unexposed groups diverge by factors other than exposure, experimental and cohort research take place. When the characteristics of the cases and the controls differ, a case-control study is conducted. It can happen in any epidemiological study. It differs from bias in that it is a fundamental aspect of the populace.

It is possible to think of confounding as the blending of effects. Because other factors linked with the exposure and the disease are also taken into account, it estimates the influence of exposure on disease is distorted. This is known as the "third variable" problem.

When a variable satisfies the following requirements, it is regarded as a confounder.

a. Independent cause or predictor of the outcome or illness: Both exposed and untainted people are affected by the confounder's link to the disease.

b. Linked to the population's exposure that gave rise to cases: The confounder is more or less prevalent in the exposed group than in the comparison group.

c. Cannot be a step in the chain of events linking exposure to disease: The exposure cannot be the source of the confounder [7].

B. Controlling methods for confounding

i. Randomization:

To ensure that each participant has an equal chance of being assigned to the treatment or comparison group, randomly assign study participants to treatment groups. Randomization ensures baseline comparability of the exposed and unexposed groups in terms of both known and unknown confounders when there are enough individuals. This only functions when the trial is sufficiently large, the investigator's impact on treatment assignment is unaffected, and when it fails, confounding must be taken into account in the analysis. There is no upper limit to the number of confounders that can be adjusted for, and there is no requirement for information about unidentified confounders, their identification, or measurement. It can only be used in experimental investigations, and smaller sample sizes make it less effective.

ii. Restriction:

It restricts the study to participants who fall into a particular confounder category. For instance, if age is a confounding factor, only people over and under 65 should be included in the study population. It has the advantages of being clear, conceptual, and useful. Effective regulation of the restricted traits. Only confounders that can be measured and known are allowed. If the restriction is too broad, confusion cannot be fully controlled. Restricts sample size, makes it difficult to evaluate constrained variables, and reduces the generalizability of conclusions.

iii. Matching:

Choose research participants so that confounders are equally distributed across exposed and unexposed groups in cohort studies or between cases and controls in case-control studies. For instance, age and sex are recognized to be confounders when analyzing the relationship between smoking and lung cancer. A 65-year-old male with lung cancer was the case in the case-control research on lung cancer, and a 65-year-old male without lung cancer served as the matched control. A 48-year-old female smoker served as the index in a cohort research on the consequences of smoking, while a 48-year-old female non-smoker served as the matched comparison.

Strengths are important for complex or hard to capture factors since they allow for the simple and efficient control of the attributes being matched. It is only feasible for known, quantified confounders; finding suitable matches can be challenging, costly, and time-consuming; and matched variables cannot be evaluated.

iv. Stratification:

In order to divide the research population into subgroups, the confounder trait must be present in one group while it is absent in the other. The measure of association is then determined for each subgroup. Strengths include simplicity and ease of execution. It provides efficient control over the tiered properties. Due to issues with insufficient data, controlling multiple confounders simultaneously is challenging. Continuous variables and difficult presentations are difficult to stratify.

A measure that has been adjusted for confounding must be compared to a crude/unadjusted measure of association in order to ascertain whether confounding has taken place.

$$Magnitude\ of\ Confounding = \frac{RR_{crude} - RR_{adjusted}}{RR_{adjusted}}\ x\ 100 \quad (15)$$

According to the 10% Rule, epidemiologists must evaluate if confounding is significant when it happens. With the change-in-estimate guideline, which typically has a cut-point of 10%, it can be determined that confounding is important when the adjusted RR differs from the crude by X% or more.

v. Multivariate Regression:

Requires building a statistical model to explain the relationship between exposure, disease, and confounders. It makes simultaneous adjustments for a number of factors. Data must fit into a statistical model that is already accessible, which is tough to understand. For example, Cox proportional hazards model for longitudinal data, multivariate logistic regression for dichotomous outcomes, and multiple linear regression for continuous outcomes.

Despite efforts to control or compensate for confounding, residual confounding still exists. The use of large categories of a confounder in analysis, such as vast age groups, smokers versus non-smokers, etc., as well as confounders for whom no data were obtained, faulty confounder data, and sources are sources.

IX. EPIDEMIOLOGICAL APPROACH TO CAUSATION

Requires building a statistical model to explain the relationship between exposure, disease, and confounders. It makes simultaneous adjustments for a number of factors. Data must fit into a statistical model that is already accessible, which is tough to understand. For example, Cox proportional hazards model for longitudinal data, multivariate logistic regression for dichotomous outcomes, and multiple linear regression for continuous outcomes.

Despite efforts to control or compensate for confounding, residual confounding still exists. The use of large categories of a confounder in analysis, such as vast age groups, smokers versus non-smokers, etc., as well as confounders for whom no data were obtained, faulty confounder data, and sources are sources. The relationship between a causal component and its effect, the temporal order in which the cause must come before the effect, which may be immediate or distant in time, and the direction in which the association between the two must be asymmetrical are the three key characteristics of a cause.

Throughout history, the germ theory of sickness, the divine retribution of sins, the imbalance of body humors brought on by air, water, land, etc., and miasma that is spread by clouds have all contributed to the development of causation theory. The most current model of causation, created in the 1960s, was the causal web. There are numerous interrelated elements that contribute to sickness and paradigm shifts. The concept of multi-factorial disease causation matches this paradigm better for non-infectious disorders.

Another multi-factorial causal model is the sufficient cause model, which conceptualizes causes as pies and pie portions. For disease to happen, all the pie parts have to line up. The "sufficient reason" refers to the entire pie. A complete causal process that invariably results in disease is referred to as a "sufficient cause." "Component causes" is the name given to the pie components. These are the main subjects of our investigation since they are contributing elements to a sufficient cause.

Each pie in SCM symbolizes a sufficient cause, which is an entire causal mechanism that leads to a certain event. A component cause is a wedge that stands in for an etiologic element. A component cause could be genetic, environmental, or behavioral. All of the sufficient causes contain component cause "A," which is why they are referred to be necessary causes.

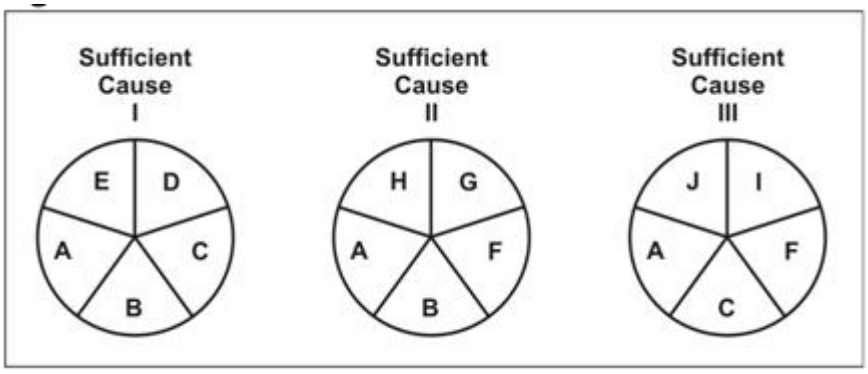

Figure 3: Causal pie model representing sufficient causes

For instance, the causes of TB infection include genetic predisposition, inadequate nutrition, crowded living conditions, poor ventilation, and, of course, exposure to the tubercle bacillus. The "necessary" cause is defined as exposure to TB. A set of standards for evaluating causation and separating association from causality was put forth by Sir Austin Bradford Hill. He proposed nine factors to help with causal inference that do not offer conclusive proof of or against causality [12].

A. Strength of Association:

Stronger relationships are more suggestive of the exposure being causal, it is asserted. When the magnitude is great, bias is less likely to account for the relationship that was seen. Example: Smoking and lung cancer. Current criticisms should not discount a tiny organization based solely on its size.

B. Consistency:

The link is more likely to be causal if it is consistently seen under various conditions. Diverse demographics, study methods, time periods, etc., could all be referred to as "different situations." Current criticisms say that study outcomes can vary for valid reasons at times. For instance, one study might have looked at low exposure levels, whereas another might have looked at high exposure levels.

C. Specificity:

It claims that, a cause, not several diseases, leads to a single disease. There is only one, not several, cause for a particular effect. Current criticisms represents stems from Koch's hypotheses about microbial disease. For non-infectious exposures, it fails. By considering smoking and lung cancer, both claims can be refuted.

D. Temporality:

The exposure must occur before the disease manifests. The disease cannot be caused by variables that occur simultaneously with the disease or factors that emerge as a result of the disease.

E. Biological Gradient:

The exposure is considered to be a cause if different exposure dosages are linked to various connections with the result. In other words, the link is more likely to be causal if it becomes stronger with increasing exposure levels. Current criticisms says that it is an old-fashioned laboratory experiment and ignores threshold effects as a reality.

F. Plausibility/Coherence:

There should be biological or social explanations for associations, it is asserted. Association should not be at odds with current understanding of the biology and natural history of diseases. Modern criticisms say that, before biological mechanisms were discovered, numerous epidemiological studies found cause-and-effect connections. An example is the discovery of carcinogenic chemicals in cigarette smoke following the early epidemiological research that established a relationship between smoking and lung cancer [2].

X. ROLE OF EPIDEMIOLOGY IN PUBLIC HEALTH

Because it describes health and disease in communities rather than in individuals, epidemiology is the foundational science of public health. This information is crucial for the development of successful public health efforts that aim to prevent illness and increase community health.

In order to uncover trends, epidemiologists use data on illness, injury, and other adverse health outcomes that they gather, analyze, and use in the field of public health. In the end, epidemiologists' work aids in improving population health by reducing and controlling adverse health consequences. Epidemiology is therefore a vital component of the study of public health and a desirable career option.

Natural history to clarify particular abnormalities in the biological system of the host and to increase diagnostic accuracy, community diagnosis to determine the morbidity and mortality from specific diseases, and the spectrum of disease that was caused by numerous agents and conditions are all useful. By describing the clinical picture of the disease, including symptoms and signs, the extent of an epidemic, risk factors, and the causative agent, and by assisting in determining treatment effectiveness and control effects, public health funds can be used effectively to control the diseases that have the greatest detrimental effects on community health.

It also addresses the identification of risk factors that either increase or decrease the likelihood of contracting a disease, the detection of disease precursors and symptoms, the evaluation of public health initiatives, the investigation of epidemics with unknown causes, and the elucidation of molecular and genetic determinants of disease progression.

In the realm of public health, epidemiologists are essential in gathering and evaluating data about illnesses, injuries, and other adverse health outcomes in order to spot trends. In the end, epidemiologists' work contributes to improving the health of their people and reducing adverse health consequences. As a result, epidemiology is an essential component of the field of public health and a rewarding career option.

XI. REFERENCES

1. Aggarwal, R., & Ranganathan, P. (2019). Study designs: Part 2–descriptive studies. *Perspectives in clinical research*, *10*(1), 34.
2. Alrawahi, A. H. (2020). New approaches to disease causation research based on the sufficient-component cause model. *Journal of Public Health Research*, *9*(3), jphr-2020.
3. Aschengrau, A., & Seage, G. R. (2013). *Essentials of epidemiology in public health*. Jones & Bartlett Publishers.
4. Friis, R. H., & Sellers, T. (2020). *Epidemiology for public health practice*. Jones & Bartlett Learning.
5. Gordis, L. (2013). *Epidemiology E-Book*. Elsevier Health Sciences.
6. Kumar, G., & Acharya, A. S. (2014). Biases in epidemiological studies: How far are we from the truth?. *Indian Journal of Medical Specialities*, *5*(1), 29-35.
7. McNamee, R. (2003). Confounding and confounders. *Occupational and environmental medicine*, *60*(3), 227-234.
8. Nganwa, D., Habtemariam, T., Tameru, B., Gerbi, G., Bogale, A., Robnett, V., & Wilson, W. (2010). Applying the epidemiologic problem oriented approach (EPOA) methodology in developing a knowledge base for the modeling of HIV/AIDS. *Ethnicity & disease*, *20*(1 Suppl 1), S1.
9. Noordzij, M., Dekker, F. W., Zoccali, C., & Jager, K. J. (2010). Measures of disease frequency: prevalence and incidence. *Nephron Clinical Practice*, *115*(1), c17-c20.
10. Rothman, K. J. (2012). *Epidemiology: an introduction*. Oxford university press.
11. Silman, A. J., Macfarlane, G. J., & Macfarlane, T. (2018). *Epidemiological studies: a practical guide*. Oxford University Press.
12. Wensink, M., Westendorp, R. G., & Baudisch, A. (2014). The causal pie model: an epidemiological method applied to evolutionary biology and ecology. *Ecology and Evolution*, *4*(10), 1924-1930.