Showing Cause, Introduction to Study Design

Principles of Epidemiology

Epidemiology

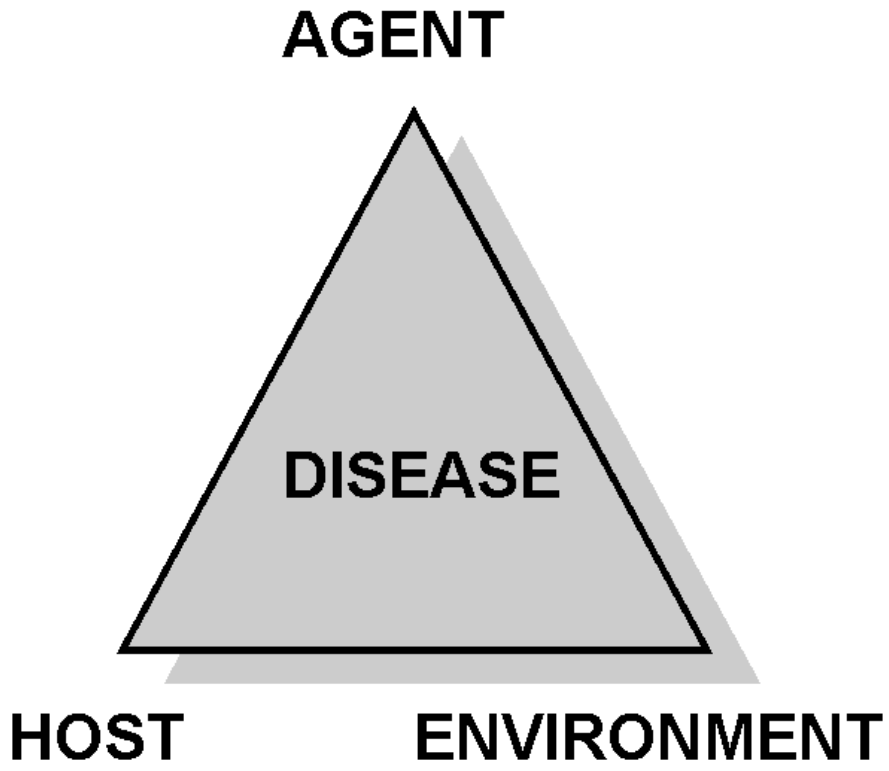
- Greek: EPI – Upon
- DEMOS – People
- LOGOS – Study of, Body of Knowledge

- **Epidemiology:** Is study of the distribution and determinants of diseases in human populations
- **Distribution**: Person, Place, Time
- **Determinants (Factors)**: Agent, Host, Environment

Epidemiologic Triad

Disease is the result of forces within a dynamic system consisting of:

- agent of infection
- host
- environment



Classic Epidemiologic Theory

- Agents
  - Living organisms
  - Exogenous chemicals
  - Genetic traits
  - Psychological factors and stress
  - Nutritive elements
  - Endogenous chemicals
  - Physical forces
- Agents have characteristics such as infectivity, pathogenicity and virulence (ability to cause serious disease)
  - They may be transmitted to hosts via vectors

Classic Epidemiologic Theory (cont.)

- **Host factors:**
  - **Immunity and immunologic response**
  - **Host behavior**
- **Environmental factors:**
  - **Physical environment (heat, cold, moisture)**
  - **Biologic environment (flora, fauna)**
  - **Social environment (economic, political, culture)**

Hill suggested that the following aspects of an association be considered in attempting to distinguish causal from non-causal associations:

- **Strength of Association – the stronger the association, the less likely the relationship is due to chance or a confounding variable.**
- **Consistency of the Observed Association – has the association been observed by different persons, in different places, circumstances, and times? (similar to the replication of laboratory experiments). The association is consistent if the results are replicated when studied in different settings and by different methods.**
- **Specificity – The criterion of specificity requires that a cause lead to a single effect, not multiple effects. Smoking is a cause of lung cancer.**
- **Temporality – the exposure of interest must precede the outcome by a period of time consistent with any proposed biologic mechanism. The cause must precede the effect in time.**
- **Biologic Gradient – there is a gradient of risk associated with the degree of exposure (dose-response relationship). Presence of a dose response curve.**

Hill's Postulates (cont)

6. **Biologic Plausibility –** there is a known or postulated mechanism by which the exposure might reasonably alter the risk of developing the disease.

7. **Coherence –** the observed data should not conflict with known facts about the natural history and biology of the disease.

8. **Experiment –** the strongest support for causation may be obtained through controlled experiments (clinical trials, intervention studies, animal experiments)

9. **Analogy –** If one drug can cause birth defects, perhaps another one can also cause birth defects. This could conceivably enhance the credibility that an association is causal.

Causal Relationships

- A causal pathway may be direct or indirect
- In direct causation, A causes B without intermediate effects
- In indirect causation, A causes B, but with intermediate effects
- In human biology, intermediate steps are virtually always present in any causal process

Association

*Association: is a Statistical dependence between two variables.

- Exposure (Risk factor, Protective factor, Predictor variable, Treatment)

- * Outcome (Disease, Event)

Measures of Effect

Measures of Effect are:

* Risk Difference (RD)
* Relative Risk (RR)
  - Risk Ration (RR)
  - Rate Ratio (RR)
* Odds Ratio (OR)

Measures of Disease Frequency

- Incidence: number of new cases of a disease / Population at risk
- Prevalence: number of existing cases (old and new) cases/Population at risk

P = I X D where:

P = Prevalence

I = Incidence

D = Duration of disease

Objectives of Epidemiologic study design

- Precision (Lack of Random Error). Reduction of random error.
- Validity (Lack of Systematic Error). Validity composed of two components:

a. Internal validity: inference for the study subjects themselves. Internal validity can be affected by the following types of biases:

1. Selection bias
2. 2. Information bias
3. 3. Confounding variable

b. External validity: inference for people outside the study population.

Strategies in the design of Epidemiologic Studies

Improving precision

Improving Validity

Study Designs

Means to assess possible causes by gathering and analyzing evidence

Types of Study Designs

- **Descriptive studies (to generate hypotheses)**
  - **Case-reports**
  - **Cross-sectional studies (prevalence studies) measure exposure and disease at the same time**
  - **Ecological studies (correlational studies) use group data rather than data on individuals**
    - **These data cannot be used to assess individual risk**
    - **To do this is to commit *ecological fallacy***

Types of Study Designs (cont.)

- Analytic studies (to test hypotheses)
  - Experimental studies
    - Clinical trials
    - Field trials
    - Intervention studies
  - Observational studies
    - Case-control studies
    - Cohort studies

The Key to Study Design

- The key to any epidemiologic study is in the definition of what constitutes a case and what constitutes exposure
- Definitions must be exclusive, categorical
- Failure to effectively define a case may lead to misclassification bias

Sources or Error in Epidemiologic Studies

- **Misclassification – wrongful classification of status for either disease or exposure**
- **Random variation - chance**

Sources or Error in Epidemiologic Studies

- **Bias** – systematic preferences built into the study design
- **Confounding** – occurs when a variable is included in the study design that is related to both the outcome of interest and the exposure, leading to false conclusions

  Example: Coffee drinking and pancreatic cancer, smoking is a confounding
- **Effect modification** – occurs when the magnitude of the association between the outcome of interest and the exposure differ according to the level of a third variable
  - The effect may be to nullify or heighten the association
  - Example: gender and hip fracture modified by age

Contingency Tables

The findings for most epidemiologic studies can be presented in the 2x2 table

	Disease		
	Yes	No	Total
Exposure			
Yes	a	b	a+b
No	c	d	c+d
Total	a+c	b+d	a+b+c+d

Measures of Association from the 2x2 Table

***Cohort Study*: the outcome measure is the *relative risk* (or risk ratio or rate ratio)**

- **In cohort studies you begin with the exposure of interest and then determine the rate of developing disease**
- **RR measures the likelihood of getting the disease if you are exposed relative to those who are unexposed**
  - **RR = incidence in the exposed/incidence in the unexposed**

$$RR = \frac{a/(a+b)}{c/(c+d)}$$

Measures of Association from the 2X2 Table

Case-control study: the outcome measure is an estimate of the relative risk or the ***odds ratio*** (relative odds)

- **In a case-control study, you begin with disease status and then estimate exposure**
  - **RR is estimated because patients are selected on disease status and we cannot calculate incidence based on exposure**
  - **The estimate is the odds ratio (OR) or the likelihood of having the exposure if you have the disease relative to those who do not have the disease**

$$\sim RR = OR = \frac{a/c}{b/d} = \frac{ad}{bc}$$

Attributable Risk or Risk Difference

- **In a *cohort study*, we may want to know the risk of disease attributable to the exposure in the exposed group, that is, the difference between the incidence of disease in the exposed and unexposed groups (excess risk)**

$$AR = a/(a+b) - c/(c+d)$$

AR = 0: No association between exposure and disease

AR > 0: Excess risk attributable to the exposure

AR < 0: The exposure carries a protective effect

Attributable Risk Percent

- In a ***cohort study***, we may want to know the proportion of the disease that could be prevented by eliminating the exposure in the exposed group **(attributable fraction or etiologic fraction)**

$$AR\% = AR/[a/(a+b)] \times 100$$

If the exposure is preventive, calculate the preventive fraction

Population Attributable Risk

- In a ***cohort study***, we may want to know the risk of disease attributable to exposure in the total study population or the difference between the incidence of disease in the total study population and that of the unexposed group

$$PAR = (a+c)/(a+b+c+d) - c/(c+d)$$

To estimate the PAR for a population beyond the study group you must know the prevalence of disease in the total population

Population Attributable Risk Percent

- In a ***cohort study***, we may want to know the proportion of the disease that could be prevented by eliminating the exposure in the entire study population

$$PAR\% = PAR/[(a+c)/(a+b+c+d)] \times 100$$

Summary of Attributable Risk Calculations

	In exposed group	In total population
Incidence attributable to exposure	$I_e - I_n$ AR	$I_p - I_n$ PAR
Proportion of incidence attributable to exposure	$\frac{I_e - I_n}{I_e} \times 100$ AR%	$\frac{I_p - I_n}{I_p} \times 100$ PAR%

Comparing Relative Risks

Age-Adjusted Death Rates per 100,000 for Male British Physicians

	Smokers	Non-smokers
Lung cancer	140	10
CHD	669	413

Source: Doll and Peto. Mortality in relation to smoking: Twenty years' observations on male British doctors. BMJ 1976;2:1525-36

Relative risk (relative risk, risk ratio) I_e/I_n: LC = 14.0; CHD = 1.6

Smokers are **14** times as likely as non-smokers to develop LC
Smokers are **1.6** times as likely as non-smokers to develop CHD

Smoking is a stronger risk factor for lung cancer than for CHD

Comparing Attributable Risks

Age-Adjusted Death Rates per 100,000 for Male British Physicians

	Smokers	Non-smokers
Lung cancer	140	10
CHD	669	413

Source: Doll and Peto. Mortality in relation to smoking: Twenty years' observations on male British doctors. BMJ 1976;2:1525-36

Attributable risk (risk difference, etiologic fraction) $I_e - I_n$:
LC = 130; CHD = 256

The excess of lung cancer attributable to smoking is **130** per 100,000 The excess of CHD attributable to smoking is **256** per 100,000

If smoking is causal, eliminating cigarettes would save more smokers from CHD than from LC

Comparing Attributable Risk Percents

Age-Adjusted Death Rates per 100,000 for Male British Physicians

	Smokers	Non-smokers
Lung cancer	140	10
CHD	669	413

Source: Doll and Peto. Mortality in relation to smoking: Twenty years' observations on male British doctors. BMJ 1976;2:1525-36

Attributable Risk % = $[(I_e-I_n)/I_e] \times 100$: **LC = 92%; CHD = 38%**

About **92%** of LC could be eliminated if the smokers in this study did not smoke

About **38%** of CHD could be eliminated if the smokers in this study did not smoke

If smoking is causal, eliminating cigarettes would save double the proportion of smokers from LC than CHD