Applied Epidemiology

Kostadin KOSTADINOV, MD, PhD

Department of Social Medicine and Public Health

Medical University-Plovdiv

Academic Year 2025-2026

Course Overview

Descriptive Epidemiology

- Person, Place, Time characteristics
- Measures of disease frequency (prevalence, incidence, rates)
- Basic Reproduction Number (R_0)

Analytical Epidemiology

- Bias, Bradford Hill criteria
- Measures of association (RR, OR, Risk Difference)
- Cohort, case-control, and cross-sectional studies

Experimental Epidemiology

- Randomized controlled trials (in depth)
- Vaccine efficacy, NNT, NNH

Descriptive Epidemiology

What is Epidemiology?

Definition: The study of the distribution and determinants of health-related states or events in specified populations, and the application of this study to prevent and control health problems.

Key Questions:

- Who is affected? (Person)
- Where does it occur? (Place)
- When does it occur? (Time)
- Why does it happen? (Determinants)
- How can we prevent it? (Control)

Descriptive Epidemiology: Overview

Purpose: To characterize the occurrence of disease by describing:

Person	Age, sex, ethnicity, occupation, socioeconomic status
Place	Geographic location, urban/rural, climate zones
Time	Secular trends, seasonal patterns, epidemics

Uses:

- Generate hypotheses about disease etiology
- Plan health services and allocate resources
- Monitor disease trends over time

Person Characteristics

Age: Most important demographic variable

- Infectious diseases: often higher in children (measles, chickenpox)
- Chronic diseases: typically increase with age (cardiovascular disease, cancer)

Sex: Biological and behavioral differences

- Males: higher rates of lung cancer, coronary heart disease
- Females: higher rates of autoimmune diseases, depression

Other factors:

- Ethnicity and race
- Occupation and socioeconomic status
- Marital status and family size

Place Characteristics

Geographic variations reveal important clues about disease etiology:

Example 1 - Infectious Disease: Malaria distribution follows the geographic range of Anopheles mosquitoes (tropical and subtropical regions)

Example 2 - Non-Infectious Disease: Gastric cancer rates are highest in Japan, Korea, and parts of South America, suggesting dietary or environmental factors

Considerations:

- Urban vs. rural differences
- Climate and environmental factors
- Access to healthcare

Time Characteristics

Secular Trends: Long-term changes over years or decades

- Decline of tuberculosis in developed countries (20th century)
- Rise of type 2 diabetes globally (past 40 years)

Seasonal Patterns:

- Influenza peaks in winter months
- Allergic rhinitis peaks in spring/summer

Epidemic Curves: Show the time course of an outbreak

- Point source: sharp rise and fall
- Continuous source: prolonged elevation
- Propagated: successive peaks

Understanding Disease Frequency

Before measuring disease, we need to understand key concepts:

Population at Risk: The group of people who can potentially develop the disease

- Example: Only women can develop ovarian cancer
- Example: Only unvaccinated individuals can develop vaccine-preventable diseases

Case Definition: Standardized criteria for identifying cases

- Must be clear, consistent, and applicable
- Includes clinical, laboratory, and epidemiological criteria

Ratio, Proportion, and Rate

Ratio: Comparison of two quantities (not necessarily related)

$$\text{Ratio} = \frac{a}{b}$$

Example: Sex ratio = Males / Females

Ratio, Proportion, and Rate

Proportion: A type of ratio where numerator is part of denominator

$$\text{Proportion} = \frac{a}{a+b}, \quad 0 \leq \text{Proportion} \leq 1$$

Example: Proportion of deaths from cancer = Cancer deaths / All deaths

Ratio, Proportion, and Rate

Rate: A proportion that includes a time dimension

$$\text{Rate} = \frac{\text{Number of events}}{\text{Population at risk} \times \text{Time period}}$$

Prevalence

Definition: The proportion of a population that has a disease at a specific point in time (or during a period) Point Prevalence:

$$\text{Point Prevalence} = \frac{\text{Number of existing cases at a point in time}}{\text{Total population at that point}}$$

Period Prevalence:

$$\text{Period Prevalence} = \frac{\text{Number of existing cases during a period}}{\text{Average population during period}}$$

Interpretation: Prevalence is a snapshot of disease burden

- Useful for: Healthcare planning, resource allocation

Prevalence: Examples

Example 1 - Non-Infectious (Diabetes): In a city of 100,000 people, 8,500 have been diagnosed with type 2 diabetes.

$$\text{Prevalence} = \frac{8500}{100000} = 0.085 = 8.5\%$$

Example 2 - Infectious (HIV): In a country with 10 million adults, 250,000 are living with HIV.

$$\text{Prevalence} = \frac{250000}{10000000} = 0.025 = 2.5\%$$

Note: Prevalence depends on both new cases and how long people live with the disease.

Incidence

Definition: The occurrence of new cases of disease in a population over a specified time period Two types of incidence measures: 1. Cumulative Incidence (Risk):

$$\text{CI} = \frac{\text{Number of new cases during time period}}{\text{Population at risk at start of period}}$$

2. Incidence Rate (Incidence Density):

$$\text{IR} = \frac{\text{Number of new cases}}{\text{Person-time at risk}}$$

Key difference: Cumulative incidence assumes complete follow-up; incidence rate accounts for varying follow-up times.

Cumulative Incidence: Example

Study: A cohort of 1,000 healthy adults followed for 5 years to study hypertension development.

- At study start: 1,000 individuals without hypertension
- After 5 years: 120 developed hypertension

$$\text{Cumulative Incidence} = \frac{120}{1000} = 0.12 = 12\%$$

Interpretation: 12% of the population developed hypertension over 5 years, or the risk of developing hypertension is 12%. Note: CI is a proportion (unitless), also called incidence proportion or attack rate (in outbreaks).

Incidence Rate: Example

Study: Following 500 healthcare workers for tuberculosis infection over 3 years.

- Total person-years of follow-up: 1,200 person-years
- New TB infections: 18

$$\text{Incidence Rate} = \frac{18}{1200} = 0.015 \text{ per person-year}$$

Or: 15 cases per 1,000 person-years

Interpretation: For every 1,000 years of observation, we expect 15 new cases of TB. Note: IR has units of 1/time (per year, per month, etc.)

Person-Time: Understanding the Concept

Person-time accumulates the total time each individual is at risk:

Person	Follow-up	Outcome
A	3 years	Disease
B	5 years	No disease
C	2 years	Lost to follow-up
D	5 years	No disease
E	4 years	Disease

Total person-years: 3 + 5 + 2 + 5 + 4 = 19 person-years

New cases: 2 $\longrightarrow$ IR = 2/19 = 0.105 per person-year

Relationship Between Prevalence and Incidence

For diseases in a steady state (stable incidence and duration):

$$\text{Prevalence} \approx \text{Incidence} \times \text{Duration}$$

Implications:

- High prevalence can result from high incidence OR long duration
- Low incidence + long duration = high prevalence (e.g., diabetes)
- High incidence + short duration = low prevalence (e.g., common cold)

Example: If incidence of diabetes is 5 per 1,000/year and duration is 20 years:

Prevalence $\approx 0.005 \times 20 = 0.10 = 10\%$

Basic Reproduction Number (R_0)

Definition: The average number of secondary infections produced by one infected individual in a completely susceptible population

$$R_0 = \beta \times c \times d$$

Where:

- β = probability of transmission per contact
- c = number of contacts per unit time
- d = duration of infectiousness

$R_0 > 1$: Epidemic will grow | $R_0 = 1$: Endemic | $R_0 < 1$: Epidemic will die out

R_0 Values for Common Diseases

Disease	R_0	Herd Immunity Threshold
Measles	12-18	92-95%
Pertussis	12-17	92-94%
Diphtheria	6-7	83-86%
Polio	5-7	80-86%
Smallpox	5-7	80-86%
COVID-19 (original)	2-3	50-67%
Influenza (seasonal)	1.5-2	33-50%
Ebola	1.5-2.5	33-60%

Effective Reproduction Number (R_t)

Definition: The average number of secondary infections at time t, accounting for immunity

$$R_t = R_0 \times S$$

Where S = proportion of population still susceptible

Example: If R_0 = 3 and 50% of population is immune:

$$R_t = 3 \times 0.5 = 1.5$$

The epidemic continues but at a slower pace.

Public health goal: Reduce R_t below 1 through vaccination or interventions.

Herd Immunity Threshold

Definition: The proportion of the population that needs to be immune to stop transmission

$$\text{HIT} = 1 - \frac{1}{R_0}$$

Example - Measles: $R_0 = 15$

$$\text{HIT} = 1 - \frac{1}{15} = 0.933 = 93.3\%$$

Implications for vaccination:

- Higher $R_0 \longrightarrow$ Higher vaccination coverage needed
- Measles requires 95% coverage due to high R_0

Mortality Measures

Crude Mortality Rate:

$$\text{CMR} = \frac{\text{Total deaths in a year}}{\text{Mid-year population}} \times 1000$$

Cause-Specific Mortality Rate:

$$\text{CSMR} = \frac{\text{Deaths from specific cause}}{\text{Mid-year population}} \times 100000$$

Case Fatality Rate (Ratio):

$$\text{CFR} = \frac{\text{Deaths from disease}}{\text{Number of cases of disease}} \times 100$$

Note: CFR measures disease severity. Ebola CFR ≈ 50%, COVID-19 CFR ≈ 1-2%

Attack Rate

Definition: The proportion of a population that develops disease during an outbreak

$$\text{Attack Rate} = \frac{\text{Number of new cases}}{\text{Population at risk}} \times 100$$

Example - Foodborne Outbreak: At a wedding, 200 guests attended. 45 developed gastroenteritis.

$$\text{Attack Rate} = \frac{45}{200} \times 100 = 22.5\%$$

Secondary Attack Rate: Cases among contacts of primary cases

$$\text{SAR} = \frac{\text{Cases among contacts}}{\text{Total susceptible contacts}} \times 100$$

Morbidity and Mortality: Summary

Measure	Formula	Unit
Point Prevalence	Existing cases / Population	Proportion
Cumulative Incidence	New cases / Population at risk	Proportion
Incidence Rate	New cases / Person-time	1/Time
Mortality Rate	Deaths / Population × 1000	Per 1,000
Case Fatality Rate	Deaths / Cases × 100	Percent
Attack Rate	New cases / At risk × 100	Percent
R_0	Secondary cases / Primary case	Ratio

Analytical Epidemiology

From Description to Analysis

Descriptive epidemiology generates hypotheses about disease causes.

Analytical epidemiology tests these hypotheses by examining associations between exposures and outcomes.

Key question: Is exposure A associated with disease B?

Study designs:

- Cohort studies (follow exposed and unexposed)
- Case-control studies (compare cases to controls)
- Cross-sectional studies (snapshot analysis)

The 2×2 Contingency Table

The foundation of analytical epidemiology:

	Disease+	Disease−	Total
Exposed	a	b	a + b
Unexposed	c	d	c + d
Total	a + c	b + d	N

This table is used to calculate all measures of association!

Risk: Definition and Calculation

Risk = Probability that an individual will develop disease during a specified time

$$\text{Risk} = \frac{\text{Number who develop disease}}{\text{Number at risk at start}}$$

From our 2×2 table:

- Risk in exposed: $R_1 = \frac{a}{a+b}$
- Risk in unexposed: $R_0 = \frac{c}{c+d}$

Note: Risk = cumulative incidence. Ranges from 0 to 1 (0% to 100%).

Relative Risk (Risk Ratio)

Definition: The ratio of disease risk in exposed compared to unexposed

$$\text{RR} = \frac{R_1}{R_0} = \frac{\frac{a}{a+b}}{\frac{c}{c+d}}$$

Interpretation:

- RR = 1: No association
- RR > 1: Exposure increases risk (harmful)
- RR < 1: Exposure decreases risk (protective)

Example: RR = 3.0 means exposed have 3× the risk compared to unexposed.

Relative Risk: Example (Cohort Study)

Study: Smoking and lung cancer - 10-year follow-up

	Lung Cancer	No Cancer	Total
Smokers	150	9,850	10,000
Non-smokers	15	9,985	10,000

$$R_{\text{smokers}} = \frac{150}{10000} = 1.5\%$$

$$R_{\text{non-smokers}} = \frac{15}{10000} = 0.15\%$$

$$\text{RR} = \frac{1.5\%}{0.15\%} = 10.0$$

Interpretation: Smokers have 10× the risk of lung cancer.

Risk Difference (Attributable Risk)

Definition: The absolute difference in risk between exposed and unexposed

$$\text{RD} = R_1 - R_0$$

From our smoking example:

$$\text{RD} = 1.5\% - 0.15\% = 1.35\%$$

Interpretation: 1.35% of smokers developed lung cancer attributable to smoking.

Public health importance: RD indicates the absolute burden of disease attributable to the exposure.

Attributable Fraction in Exposed

Definition: Proportion of disease in exposed attributable to the exposure

$$\text{AF}_e = \frac{R_1 - R_0}{R_1} = \frac{\text{RR} - 1}{\text{RR}}$$

From our smoking example:

$$\text{AF}_e = \frac{10 - 1}{10} = 0.9 = 90\%$$

Interpretation: 90% of lung cancer cases in smokers are attributable to smoking.

Use: Estimates proportion of disease preventable if exposure eliminated.

Population Attributable Fraction (PAF)

Definition: Proportion of disease in the total population attributable to exposure

$$\text{PAF} = \frac{p \times (\text{RR} - 1)}{1 + p \times (\text{RR} - 1)}$$

Where p = prevalence of exposure in population

Example: 30% smokers in population, RR = 10

$$\text{PAF} = \frac{0.30 \times 9}{1 + 0.30 \times 9} = \frac{2.7}{3.7} = 73\%$$

Interpretation: 73% of lung cancer in this population could be prevented by eliminating smoking.

Odds: Definition

Odds = Ratio of probability that event occurs to probability it does not

$$\text{Odds} = \frac{p}{1-p}$$

Example: If probability of disease is 0.20 (20%):

$$\text{Odds} = \frac{0.20}{0.80} = 0.25 = 1{:}4$$

Interpretation: For every 1 person with disease, there are 4 without.

Note: Odds ≈ Risk when disease is rare, but diverge when common.

Odds vs. Risk: Comparison

Probability (p)	Risk	Odds
0.01 (1%)	0.01	0.0101
0.10 (10%)	0.10	0.111
0.20 (20%)	0.20	0.25
0.50 (50%)	0.50	1.00
0.80 (80%)	0.80	4.00

Key point: When disease is rare (< 10%), odds ≈ risk. This is the rare disease assumption.

Odds Ratio (OR)

Definition: The ratio of odds of disease in exposed vs. unexposed

$$\text{OR} = \frac{\text{Odds in exposed}}{\text{Odds in unexposed}} = \frac{a \times d}{b \times c}$$

Interpretation:

- OR = 1: No association
- OR > 1: Exposure increases odds of disease
- OR < 1: Exposure decreases odds of disease

Critical point: OR is the primary measure in case-control studies where we cannot calculate risk directly.

What is Bias?

Definition: Systematic error in design, conduct, or analysis that results in incorrect estimates

Key points:

- Bias is systematic, not random
- Leads to incorrect conclusions
- Cannot be corrected by increasing sample size
- Must be prevented in study design
- Can bias toward or away from null

Major categories: Selection bias, Information bias, Confounding

Classification of Bias

Type	Description
Selection Bias	Systematic differences in who is included
Information Bias	Systematic errors in measurement
Confounding	Third variable distorts the relationship

Each type has multiple subtypes with specific mechanisms.

Selection Bias: Definition

Definition: Relationship between exposure and outcome differs for participants vs. non-participants

When it occurs:

- Selection of study participants
- Loss to follow-up
- Non-response
- Survival patterns

Result: Study sample not representative of target population

Selection Bias: Types (Part 1)

1. Non-response bias:
- Responders differ from non-responders
- Example: Health-conscious more likely to respond to surveys

2. Volunteer bias (Self-selection):
- Volunteers differ from general population
- Example: Clinical trial participants often healthier

3. Berkson's bias (Admission rate bias):
- Hospital controls may have conditions related to exposure
- Example: Studying coffee-heart disease in hospital patients

Selection Bias: Types (Part 2)

4. Healthy worker effect:
- Employed workers healthier than general population
- Occupational studies underestimate hazards
- Example: Chemical plant workers show lower mortality

5. Survival bias (Prevalence-incidence):
- Only survivors included in prevalence studies
- Miss rapidly fatal cases

6. Loss to follow-up:
- If losses differ by exposure/outcome status
- Example: Sicker patients drop out of trials

Selection Bias: Example

Study: Case-control of cell phones and brain cancer

Problem: Controls from hospital (Berkson's bias)

	Cases	Hospital Controls
Heavy phone use	40%	35%
Light/no use	60%	65%

Issue: Hospital controls may have conditions related to phone use (accidents while texting), biasing OR toward null.

Solution: Population-based controls

Information Bias: Definition

Definition: Systematic error in measurement of exposure, outcome, or other variables

Also called: Measurement bias, Misclassification

Two types:

- Non-differential: Errors similar across groups
- Differential: Errors differ between groups

Misclassification: Non-differential

Definition: Exposure misclassification same in diseased and non-diseased

Effect: Usually biases toward null (OR/RR $\longrightarrow$ 1.0)

- Dilutes true associations
- Harder to detect real effects

Example: Self-reported diet for diet-cancer study

- Everyone has similar difficulty remembering
- Errors equally likely in cases and controls
- True OR of 2.0 might appear as 1.5

Misclassification: Differential

Definition: Misclassification differs between groups

Effect: Can bias in either direction

- More dangerous than non-differential
- Unpredictable direction

Common causes:

- Recall bias
- Interviewer bias
- Detection bias

Information Bias: Recall Bias

Definition: Cases remember exposures differently than controls

Common in: Case-control studies

Example: Birth defects and medication during pregnancy

- Mothers of affected babies remember medications better
- Controls may forget similar exposures
- Results in inflated OR

Prevention:

- Use medical records
- Prospective studies
- Blind participants to hypothesis

Information Bias: Interviewer Bias

Definition: Interviewers probe differently based on disease status

Example: Occupational exposures and lung cancer

- Interviewer knows subject is cancer patient
- Probes more deeply about chemical exposures
- Control interviews less thorough

Prevention:

- Blind interviewers to disease status
- Standardized questionnaires
- Self-administered surveys

Information Bias: Detection Bias

Definition: Outcome detected differently in exposed vs. unexposed

Example: HRT and breast cancer

- Women on HRT have more medical visits
- More mammograms performed
- Cancer detected more often
- May overestimate HRT-cancer association

Prevention:

- Standardize surveillance across groups
- Objective outcome measures
- Blind outcome assessors

Information Bias: Hawthorne Effect

Definition: Participants change behavior because observed

Example: Study of handwashing compliance

- Healthcare workers wash hands more when observed
- True compliance may be lower

Also called: Observer effect

Prevention:

- Covert observation (if ethical)
- Prolonged observation
- Unobtrusive measures

Confounding: Recap

Definition: Third variable distorts exposure-outcome association

Criteria for confounder:

1. Associated with the exposure
2. Independently causes the outcome
3. NOT on the causal pathway

Example: Alcohol and lung cancer

- Apparent association exists
- Confounded by smoking
- After controlling for smoking, association weakens

Bias Summary Table

Bias Type	Example	Prevention
Non-response	Healthy respond more	Follow-up non-responders
Healthy worker	Workers healthier	Internal comparisons
Survival	Only survivors	Use incident cases
Recall	Cases remember better	Records; prospective
Interviewer	Probe cases more	Blind interviewers
Detection	More screening in exposed	Standardize surveillance
Misclassification	Measurement errors	Validate instruments

Preventing Bias: Study Design

Selection bias:

- Clear, objective inclusion criteria
- High participation and follow-up rates
- Population-based samples

Information bias:

- Validated measurement tools
- Blind assessors
- Standardized procedures
- Objective measures when possible

Confounding:

- Randomization, restriction, matching
- Multivariable analysis

Bias vs. Confounding vs. Random Error

Feature	Bias	Confounding	Random Error
Type	Systematic	Systematic	Random
Direction	Predictable	Assessable	Unpredictable
Larger N	No change	No change	Decreases
Control	Prevention	Design/Analysis	Larger sample

Association vs. Causation

Association: A statistical relationship between exposure and outcome

Causation: The exposure actually produces the outcome

"Correlation does not imply causation"

The challenge: How do we move from observing an association to concluding that a causal relationship exists?

Why Causality Matters

Public health decisions require causal knowledge:

- Should we recommend reducing fat intake to prevent heart disease?
- Should we ban a chemical suspected of causing cancer?
- Should we implement a vaccination program?
- Should we screen for a disease?

If we act on associations that are not causal:

- Waste resources on ineffective interventions
- May cause harm (e.g., unnecessary treatment)
- Miss the true cause of disease

Historical Context

1965: Sir Austin Bradford Hill published "The Environment and Disease: Association or Causation?"

Context: Debate over whether smoking causes lung cancer

- Strong association observed in multiple studies
- Tobacco industry argued it was just correlation
- No RCT possible (unethical to randomize people to smoke)

Bradford Hill's contribution: A framework of 9 criteria to evaluate whether an observed association is likely causal

The 9 Bradford Hill Criteria

1	Strength of association
2	Consistency
3	Specificity
4	Temporality
5	Biological gradient (dose-response)
6	Plausibility
7	Coherence
8	Experiment
9	Analogy

Note: These are guidelines, not a checklist. Only temporality is absolutely required.

1. Strength of Association

Definition: The magnitude of the association (RR, OR)

Principle: Stronger associations are more likely to be causal

- Harder to explain by bias or confounding alone
- Weak associations more susceptible to hidden biases

Examples:

- Smoking and lung cancer: RR ≈ 10-20 (strong)
- Smoking and heart disease: RR ≈ 2 (moderate)
- Both are causal, but lung cancer association is more convincing

Caution: Weak associations can still be causal (e.g., secondhand smoke)

2. Consistency

Definition: The association is observed repeatedly in different:

- Populations
- Settings
- Time periods
- Study designs
- Researchers

Example - Smoking and lung cancer:

- Observed in men and women
- In multiple countries
- In cohort and case-control studies
- By different research teams

Caution: Lack of consistency may reflect different conditions, not lack of causality

3. Specificity

Definition: The exposure leads to only one disease, or the disease has only one cause

Example: Mesothelioma is almost exclusively caused by asbestos exposure

Limitations:

- Most exposures cause multiple diseases (smoking → lung cancer, heart disease, stroke)
- Most diseases have multiple causes
- This is the weakest criterion

Modern view: Specificity adds weight when present, but absence doesn't argue against causality

4. Temporality

Definition: The exposure must precede the outcome

This is the ONLY absolutely essential criterion

The cause must come before the effect

Challenges:

- Cross-sectional studies cannot establish temporality
- Long latency periods make timing difficult
- Reverse causation: Does the disease cause the exposure?

Example: Does depression cause unemployment, or unemployment cause depression?

5. Biological Gradient (Dose-Response)

Definition: Greater exposure leads to greater risk of disease

Example - Smoking and lung cancer:

Cigarettes/day	RR
Non-smoker	1.0
1-14	8
15-24	13
25+	25

Caution:

- Threshold effects exist (no effect until certain dose)
- U-shaped or J-shaped curves possible (e.g., alcohol and heart disease)

6. Plausibility

Definition: The association is biologically plausible given current knowledge

Example - Smoking and lung cancer:

- Cigarette smoke contains known carcinogens
- Tar deposits in lungs of smokers
- Carcinogens cause DNA mutations
- Mutations lead to uncontrolled cell growth

Limitations:

- Depends on current state of knowledge
- Many causal relationships were discovered before mechanism known
- “Absence of evidence is not evidence of absence”

7. Coherence

Definition: The association is consistent with:

- Natural history of the disease
- Other known facts about the disease
- Laboratory findings

Example - Smoking and lung cancer:

- Lung cancer rates rose after smoking became popular (with lag)
- Lung cancer rates higher in countries with more smoking
- Cell changes seen in smokers' lungs match carcinogenesis pathway

Difference from plausibility: Coherence is about fitting with epidemiological and pathological knowledge, not just biological mechanisms

8. Experiment

Definition: Experimental evidence supports the association Types of experimental evidence:

- Randomized controlled trials in humans
- Natural experiments
- Removal of exposure reduces disease
- Animal experiments

Examples:

- Smoking cessation reduces lung cancer risk
- Removal of contaminated water pump stopped cholera (John Snow)
- HPV vaccine reduces cervical cancer

Strongest evidence but often not possible or ethical

9. Analogy

Definition: Similar exposures cause similar diseases

Examples:

- If thalidomide causes birth defects, other drugs might too
- If one virus causes cancer (HPV), others might (HBV, EBV)
- If one occupational dust causes lung disease, similar dusts might

Use: Helps in accepting new causal hypotheses based on established ones

Limitations: Weakest criterion - similar exposures don’t always have similar effects

Case Study: Smoking and Lung Cancer

Criterion	Met?	Evidence
1. Strength	✓	RR = 10-20 for heavy smokers
2. Consistency	✓	Hundreds of studies, multiple countries
3. Specificity	Partial	Smoking causes many diseases
4. Temporality	✓	Smoking precedes cancer by decades
5. Dose-response	✓	More cigarettes $\longrightarrow$ higher risk
6. Plausibility	✓	Carcinogens in smoke damage DNA
7. Coherence	✓	Rates parallel smoking trends
8. Experiment	✓	Cessation reduces risk; animal studies
9. Analogy	✓	Other inhaled carcinogens cause cancer

Case Study: H. pylori and Peptic Ulcer

Criterion	Met?	Evidence
1. Strength	✓	OR > 4 for ulcer in infected
2. Consistency	✓	Global studies confirm
3. Specificity	Partial	Also causes gastritis, cancer
4. Temporality	✓	Infection precedes ulcer
5. Dose-response	Partial	Bacterial load correlates
6. Plausibility	✓	Bacteria damage gastric mucosa
7. Coherence	✓	Fits with ulcer pathophysiology
8. Experiment	✓	Antibiotics cure ulcers
9. Analogy	✓	Other bacteria cause GI disease

Nobel Prize 2005: Marshall & Warren

Using the Criteria: Guidelines

The criteria are NOT:

- A checklist (don't count how many are met)
- Equally weighted
- All required (except temporality)
- A mathematical formula

The criteria ARE:

- A framework for thinking about causality
- Useful for weighing evidence
- Guidelines that require judgment
- Context-dependent

Key insight: Consider all evidence together and use scientific judgment

Hierarchy of Criteria

Essential:

Temporality

Strong evidence:

Strength | Dose-response | Experiment

Supporting evidence:

Consistency | Plausibility | Coherence

Weak criteria:

Specificity | Analogy

Modern Perspectives

Additions since 1965:

- Reversibility: Removing exposure reduces disease risk
- Study quality: Well-designed studies carry more weight
- Effect modification: Understanding who is affected and why

Causal inference methods:

- Directed Acyclic Graphs (DAGs)
- Mendelian randomization
- Instrumental variables
- Propensity score methods

Bradford Hill criteria remain foundational framework

Common Pitfalls

1. Treating criteria as a checklist
- Wrong: “6 out of 9 criteria met, so it’s causal”
- Right: Weigh all evidence with judgment

2. Ignoring alternative explanations
- Always consider: Bias? Confounding? Chance?

3. Requiring all criteria
- Only temporality is essential
- Many proven causes don’t meet all criteria

4. Overweighting specificity
- Most exposures cause multiple outcomes
- Most diseases have multiple causes

“All scientific work is incomplete… That does not confer upon us a freedom to ignore the knowledge we already have, or to postpone the action that it appears to demand at a given time.”

— Sir Austin Bradford Hill, 1965

Cross-Sectional Studies: Design

Concept: Measure exposure and outcome at the same point in time

Characteristics:

- No follow-up period
- Measures prevalence, not incidence
- Exposure and outcome assessed simultaneously
- Cannot establish temporal sequence

Measure of association: Prevalence Ratio (PR) or Prevalence OR

$$\text{PR} = \frac{\text{Prevalence in exposed}}{\text{Prevalence in unexposed}}$$

Cross-Sectional Study: Example

Study: Association between sedentary work and back pain

	Back Pain	No Pain	Total
Sedentary job	120	280	400
Active job	60	340	400

$$\text{Prevalence}_{\text{sedentary}} = \frac{120}{400} = 30\%$$

$$\text{Prevalence}_{\text{active}} = \frac{60}{400} = 15\%$$

$$\text{PR} = \frac{30\%}{15\%} = 2.0$$

But which came first - the job or the pain?

Cross-Sectional Studies: Pros and Cons

Advantages	Disadvantages
Quick and inexpensive	Cannot establish causation
No loss to follow-up	Temporal ambiguity
Multiple exposures/outcomes	Prevalence, not incidence
Good for prevalence surveys	Survival bias possible
Generate hypotheses	Not for rare diseases
Useful for health planning	Recall bias for exposures

Case-Control Studies: Design

Concept: Start with outcome (cases/controls), look back at exposure

Steps:

1. Identify cases (people with disease)
2. Select controls (disease-free from same population)
3. Determine past exposure in both groups
4. Compare exposure odds

Key features:

- Efficient for rare diseases
- Can study multiple exposures
- Cannot calculate incidence or risk directly
- OR estimates RR when disease is rare

Case-Control: Example (Infectious)

Study: Risk factors for severe COVID-19

	Severe	Mild	Total
Diabetes	80	40	120
No Diabetes	120	160	280
Total	200	200	400

$$\text{OR} = \frac{80 \times 160}{40 \times 120} = \frac{12800}{4800} = 2.67$$

Interpretation: Odds of severe COVID are 2.67× higher with diabetes.

Case-Control Studies: Pros and Cons

Advantages	Disadvantages
Efficient for rare diseases	Cannot calculate incidence
Quick and inexpensive	Recall bias
Study multiple exposures	Selection bias in controls
Small sample size needed	Temporal sequence unclear
Good for long latency	Cannot study rare exposures
Outbreak investigations	Only one outcome studied

Cohort Studies: Design

Concept: Start with exposure status, follow forward to outcome

Types:

- Prospective: Exposure assessed now, follow into future
- Retrospective: Use historical records

Key features:

- Can calculate incidence, risk, and RR directly
- Establishes temporal sequence
- Can study multiple outcomes
- Time-consuming and expensive for rare outcomes

Cohort Study: Example (Non-Infectious)

Study: Obesity and Type 2 Diabetes (5-year follow-up)

	Diabetes	No Diabetes	Total
Obese	240	1,760	2,000
Not Obese	80	2,920	3,000

$$R_{\text{obese}} = \frac{240}{2000} = 12\%$$

$$R_{\text{not obese}} = \frac{80}{3000} = 2.7\%$$

$$\text{RR} = \frac{12\%}{2.7\%} = 4.5$$

Interpretation: Obese individuals have 4.5× the risk of diabetes.

Cohort Studies: Pros and Cons

Advantages	Disadvantages
Calculate incidence, RR	Expensive and time-consuming
Establishes temporality	Loss to follow-up
Study multiple outcomes	Not efficient for rare diseases
Less recall bias	Exposure may change over time
Study rare exposures	Large sample size needed
Direct risk measurement	Long wait for results

Comparing Study Designs

Feature	Cross-sectional	Case-control	Cohort
Direction	Snapshot	Outcome⟶Exposure	Exposure⟶Outcome
Time	One point	Retrospective	Prospective/Retro
Measure	Prevalence	Odds Ratio	RR, Incidence
Best for	Surveys	Rare diseases	Common diseases
Temporality	No	Weak	Yes
Cost	Low	Low-Medium	High

When OR Approximates RR

The OR approximates RR when disease is rare (< 10%):

Mathematical basis:

When disease is rare: $a \ll b$ and $c \ll d$

So: $a + b \approx b$ and $c + d \approx d$

$$\text{RR} = \frac{\frac{a}{a+b}}{\frac{c}{c+d}} \approx \frac{\frac{a}{b}}{\frac{c}{d}} = \frac{ad}{bc} = \text{OR}$$

Caution: For common diseases, OR overestimates RR when OR > 1.

Confidence Intervals

Purpose: Quantify precision and statistical significance

95% CI: Range where true population value likely lies

For OR and RR:

- 95% CI includes 1.0 $\longrightarrow$ Not statistically significant ($p \geq 0.05$)
- 95% CI excludes 1.0 $\longrightarrow$ Statistically significant ($p < 0.05$)

Example: OR = 2.5 (95% CI: 1.8 - 3.5)

The CI excludes 1.0, so the association is statistically significant.

Confounding

Definition: A third variable distorts the exposure-outcome relationship

Confounder criteria:

1. Associated with the exposure
2. Independently causes the outcome
3. NOT on the causal pathway

Example: Coffee → Heart disease?

- Confounded by smoking
- Smokers drink more coffee AND smoking causes heart disease
- Apparent association may be due to smoking

Controlling for Confounding

Study design:

- Randomization (experiments)
- Restriction (limit to one stratum)
- Matching (pair exposed and unexposed)

Analysis:

- Stratification (analyze subgroups separately)
- Multivariable adjustment (regression)

Mantel-Haenszel OR:

$$\text{OR}_{\text{MH}} = \frac{\sum a_i \frac{d_i}{n_i}}{\sum b_i \frac{c_i}{n_i}}$$

Effect Modification (Interaction)

Definition: Association magnitude differs across subgroups

Example: Oral contraceptives and stroke

- In non-smokers: OR = 2.0
- In smokers: OR = 8.0

Smoking modifies the effect of OCs on stroke risk.

Difference from confounding:

- Confounding: Control for it (hide it)
- Effect modification: Report it (reveal it)

Report stratum-specific measures when present!

Experimental Epidemiology

Experimental vs. Observational Studies

Observational studies:

- Researcher observes without intervention
- Cannot prove causation (association only)
- Subject to confounding

Experimental studies:

- Researcher assigns the exposure/intervention
- Random allocation eliminates confounding
- Strongest evidence for causation

Types: RCTs, Field trials, Community trials

Field Trials

Definition: Experiments in community settings, not clinical

Characteristics:

- Participants are healthy individuals
- Interventions often preventive
- Larger samples needed (lower event rates)
- Real-world conditions

Examples:

- Vaccine trials in endemic areas
- Vitamin supplementation trials
- Fluoride trials for dental caries

Community Intervention Trials

Definition: Unit of randomization is a community/group

Also called: Cluster randomized trials

Examples:

- Water fluoridation (cities randomized)
- School-based health education
- Workplace smoking cessation
- Village mosquito net distribution

Challenges:

- Fewer units → less power
- Outcomes correlated within clusters
- Contamination between groups

Phases of Clinical Trials

Phase I: Safety and dosing (20-100 healthy volunteers)

- Identify safe dose range and side effects

Phase II: Efficacy and safety (100-300 patients)

- Does the treatment work?

Phase III: Confirmatory (1,000-3,000+ patients)

- Confirm efficacy, monitor adverse events
- Required for regulatory approval

Phase IV: Post-marketing surveillance

- Ongoing safety monitoring
- Detect rare adverse events

Ethical Considerations in RCTs

Fundamental principles:

1. Equipoise: Genuine uncertainty about best treatment
2. Informed consent: Fully informed, voluntary
3. Beneficence: Maximize benefit, minimize harm
4. Justice: Fair selection of participants

Requirements:

- Ethics committee approval (IRB/IEC)
- Data Safety Monitoring Board (DSMB)
- Stopping rules for efficacy or harm

CONSORT Guidelines

Consolidated Standards of Reporting Trials

Purpose: Standardize RCT reporting for transparency

Key elements:

- Participant flow diagram
- Randomization method
- Blinding procedures
- Primary and secondary outcomes
- Sample size calculation
- Intention-to-treat analysis
- Adverse events

Randomized Controlled Trials: Overview

The gold standard for testing interventions

Key features:

1. Random allocation: Participants randomly assigned
2. Control group: Comparison with placebo/standard care
3. Blinding: Participants/researchers unaware of assignment
4. Prospective: Follow-up from randomization to outcome

Why randomization matters:

- Balances known AND unknown confounders
- Eliminates selection bias
- Enables causal inference

RCT: Step-by-Step Process

1. Define research question (PICO):
- P: Population
- I: Intervention
- C: Comparator
- O: Outcome

2. Sample size calculation:
- Based on effect size, α (0.05), power (80%)

3. Recruitment:
- Inclusion/exclusion criteria
- Informed consent

RCT: Randomization Methods

Simple randomization:

- Coin flip or random numbers
- Risk of imbalance in small trials

Block randomization:

- Equal group sizes within blocks
- Blocks of 4: AABB, ABAB, ABBA, etc.

Stratified randomization:

- Randomize within strata (age, severity)
- Ensures balance on key factors

Cluster randomization:

- Groups (clinics, schools) randomized

RCT: Allocation Concealment

Definition: Preventing enrollment staff from knowing upcoming assignment

Why important:

- Prevents selection bias
- Enrollers cannot cherry-pick patients

Methods:

- Central randomization (phone/web)
- Sequentially numbered opaque sealed envelopes
- Pharmacy-controlled allocation

Note: Allocation concealment ≠ Blinding

Types of Blinding

Single-blind:

- Participants do not know assignment
- Reduces placebo effect

Double-blind:

- Neither participants nor investigators know
- Reduces measurement bias

Triple-blind:

- Participants, investigators, AND analysts blinded
- Maximum bias reduction

Open-label: No blinding (sometimes necessary)

RCT: Control Groups

Placebo control:

- Inert substance/sham procedure
- Best for measuring true effect
- Only ethical when no proven treatment exists

Active control:

- Comparison with existing standard treatment
- When placebo would be unethical

Usual care:

- Whatever treatment normally received

Waitlist control:

- Controls receive intervention after study

RCT: Outcome Assessment

Primary outcome:

- Main outcome study designed to detect
- Usually one, pre-specified
- Sample size based on this

Secondary outcomes:

- Additional outcomes of interest
- Hypothesis-generating

Types:

- Hard endpoints: Death, stroke (objective)
- Surrogate: BP, cholesterol (may not reflect clinical benefit)
- Patient-reported: QoL, pain scores

RCT: Analysis Approaches

Intention-to-Treat (ITT):

- Analyze ALL as randomized
- Regardless of compliance
- Preserves randomization
- Conservative estimate

Per-Protocol (PP):

- Only those who completed treatment
- May overestimate effect
- Subject to bias

Primary analysis should always be ITT!

Vaccine Efficacy (VE)

Definition: Proportional reduction in disease risk among vaccinated

$$\text{VE} = (1 - \text{RR}) \times 100 = \frac{R_{\text{unvax}} - R_{\text{vax}}}{R_{\text{unvax}}} \times 100$$

Interpretation:

- VE = 95% means 95% fewer cases in vaccinated group
- VE = 0% means no protection
- VE < 0% indicates increased risk (rare)

Note: Efficacy (controlled trials) vs. Effectiveness (real-world)

Vaccine Efficacy: Example

Study: COVID-19 vaccine trial (30,000 participants)

	COVID+	COVID−	Total
Vaccine	8	14,992	15,000
Placebo	162	14,838	15,000

$$R_{\text{vaccine}} = \frac{8}{15000} = 0.00053$$

$$R_{\text{placebo}} = \frac{162}{15000} = 0.0108$$

$$\text{RR} = \frac{0.00053}{0.0108} = 0.049$$

$$\text{VE} = (1 - 0.049) \times 100 = 95.1\%$$

Vaccine Efficacy: More Examples

Vaccine	VE	Outcome
Measles (2 doses)	97%	Measles infection
HPV (Gardasil 9)	90%	HPV-related cancers
Influenza	40-60%	Influenza illness
Pneumococcal	75%	Invasive disease
Rotavirus	85-98%	Severe gastroenteritis
COVID-19 mRNA	94-95%	Symptomatic COVID

Note: VE varies by outcome (infection vs. severe disease vs. death)

Measures in Clinical Trials

Relative Risk Reduction (RRR):

$$\text{RRR} = \frac{R_0 - R_1}{R_0} = 1 - \text{RR}$$

Absolute Risk Reduction (ARR):

$$\text{ARR} = R_0 - R_1$$

Number Needed to Treat (NNT):

$$\text{NNT} = \frac{1}{\text{ARR}}$$

Interpretation: NNT = patients needed to treat for 1 to benefit

NNT Example: Cardiovascular Prevention

Study: Statin therapy for heart attack prevention (5 years)

- Control group rate: 8% ($R_0 = 0.08$)
- Statin group rate: 5% ($R_1 = 0.05$)

$$\text{RRR} = \frac{0.08 - 0.05}{0.08} = 37.5\%$$

$$\text{ARR} = 0.08 - 0.05 = 0.03 = 3\%$$

$$\text{NNT} = \frac{1}{0.03} = 33.3 \approx 34$$

Interpretation: Treat 34 patients for 5 years to prevent 1 heart attack.

Number Needed to Harm (NNH)

Definition: Patients to treat for 1 additional adverse event

$$\text{NNH} = \frac{1}{\text{ARI}}$$

Where ARI = Absolute Risk Increase

Example: Antibiotic causing diarrhea

- Antibiotic group: 15% diarrhea
- Placebo group: 5% diarrhea
- ARI = 0.15 - 0.05 = 0.10
- NNH = 1/0.10 = 10

Interpretation: For every 10 treated, 1 additional gets diarrhea.

NNT vs. NNH: Clinical Decisions

Scenario	NNT	NNH	Decision
Statin prevention	34	250	Favorable
Aspirin for MI	100	300	Favorable
Aggressive chemo	5	3	Discuss with patient
New antibiotic	50	10	Consider alternatives

Rule: NNT << NNH is favorable. Consider severity of outcomes!

Thank You

Questions?

Kostadin Kostadinov, MD, PhD

Department of Social Medicine and Public Health

Medical University-Plovdiv