

Biomedical
Science

Molecular
Genetics



BMS 23010c | Fall 2026-2027

Epidemiology & Public Genetics
Introduction to Genetic Epidemiology



Free Session

A+ is more than just a platform



[Academy Website](#)



[WhatsApp Community](#)



What is Epidemiology?

Distribution

Study of how diseases
distributed
in populations

Determinants

Identifies
factors that influence
health

Goal

Prevent
and
treat
diseases



Major Goals in Epidemiology

1

To obtain an **unbiased & precise estimate** of the true effect of an exposure or intervention on outcome in the **population at risk**.

2

To use this knowledge to **prevent and treat disease**.



Core Concepts: Incidence & Prevalence

Incidence

New cases in a time period.

183,529 new prostate cancer cases (US, 2015)

Prevalence

All existing cases at a given time.

3.1 million men living with prostate cancer (US, 2015)

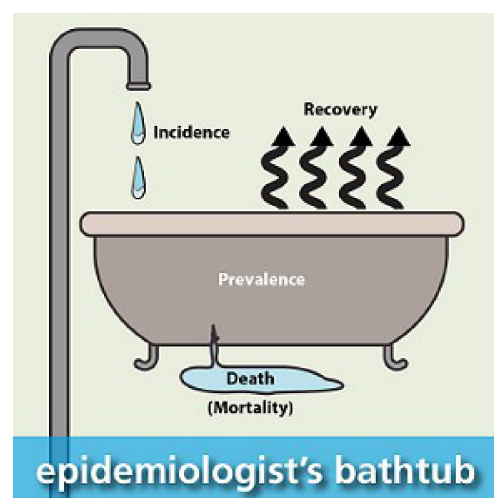
$$\text{Prevalence} = \frac{\text{All current cases (new + old)}}{\text{Total population}}$$

THE EPIDEMIOLOGIST'S BATHTUB

Incidence = tap filling the tank (new cases)

Recovery & Death = drains emptying the tank (cases leaving)

Prevalence = water currently in the tank (active cases)





Study Designs

Cohort Study → follow healthy people over time

Relative Risk (also called Risk Ratio) is used in cohort studies. It compares the **risk of developing disease** in the exposed group with the risk in the unexposed group.

Case-Control Study → compare sick (cases) vs healthy (controls)

Odds Ratio (OR) is a measure used in case-control studies to see if an exposure is associated with a disease. It compares the **odds of exposure among cases** (diseased) to the odds of exposure among controls (healthy).

Cohort



Relative Risk (RR)

Case-Control

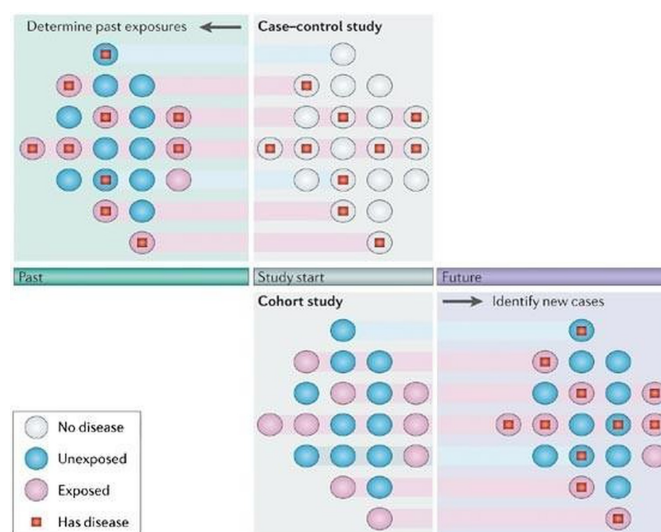


Odds Ratio (OR)



Case-Control vs. Cohort Studies

FEATURE	CASE-CONTROL STUDY	COHORT STUDY
Starting point	Begin with diseased (cases) and healthy (controls)	Begin with healthy people (some exposed, some not)
Direction	Looks backward to check past exposure	Looks forward to see who develops disease
Time	Usually retrospective	Usually prospective (can also be retrospective)
Cost & Time	Cheaper, quicker	More expensive, longer follow-up
Best for	Rare diseases, long latency diseases	Common diseases, studying multiple outcomes
Measure	Odds Ratio (OR)	Relative Risk (RR)
Weakness	Recall bias, cannot directly measure incidence	Loss to follow-up, costly



Copyright © 2006 Nature Publishing Group
Nature Reviews | Genetics

Manolio et al. Nature Reviews Genetics 7, 812–820 (October 2006) | doi:10.1038/nrg1919



Association and Causality

An exposure and outcome are **associated** if there is a differential distribution:

- **Incidence of outcome** differs for exposed and unexposed group (cohorts); or
- **Prevalence of exposure** differs between cases and controls (case-control study)

An exposure is **causal** for the outcome if the presence (or absence) of the exposure directly or indirectly influences whether the outcome occurs.

Association

Two things occur together (e.g., smoking & lung cancer)

Causality

One directly influences the other

Reminder: Association $\neq$ Causation



Challenges: Bias

DEFINITION

A **systematic error** in how a study is designed, conducted, or analyzed that leads to incorrect results. It makes the results **consistently wrong in one direction**.

Selection Bias

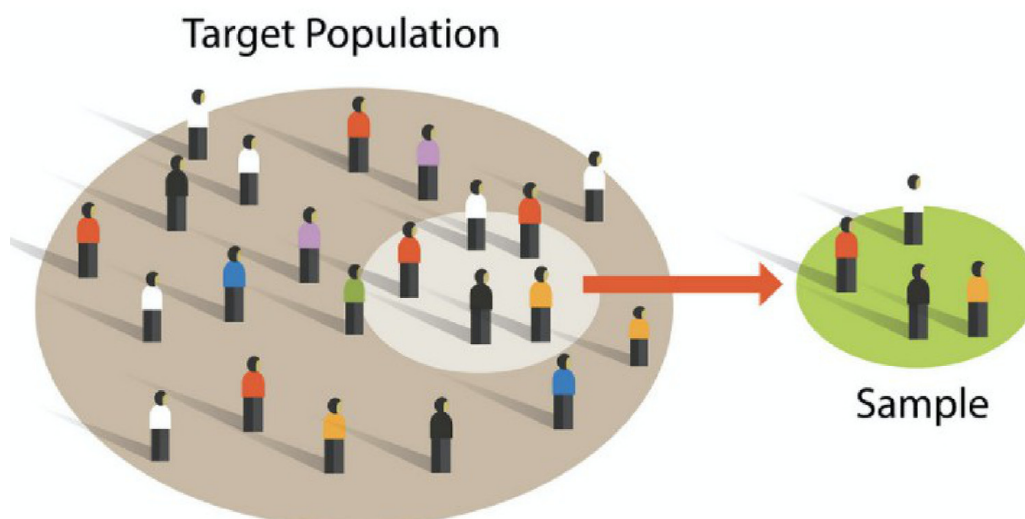
When participants chosen for the study are **not representative** of the population.

Example: Studying smoking and lung cancer but recruiting only hospital patients. Smokers may be overrepresented.

Recall Bias

In case-control studies, **cases may remember exposures better** than controls.

Example: Mothers of babies with birth defects may recall medication use more clearly than mothers of healthy babies.

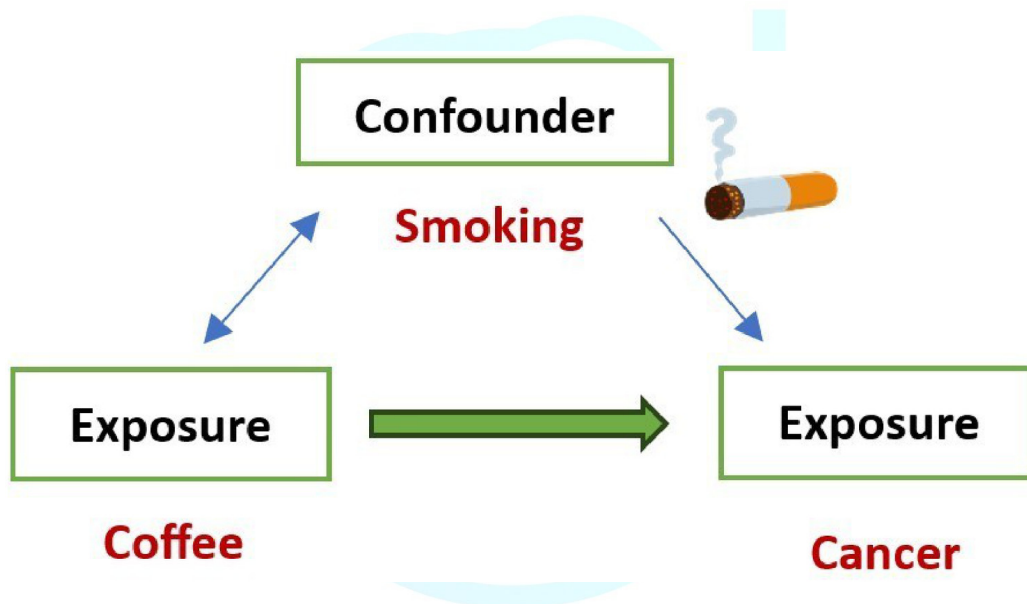




Challenges: Confounding

DEFINITION

When a **third factor** (the “confounder”) distorts the true relationship between exposure and outcome. It makes it look like there is an association (or hides one), but the effect is really due to **another variable**.



WORKED EXAMPLE

- 1 Study:** Does coffee drinking cause lung cancer?
- 2 Finding:** Coffee drinkers seem to have higher lung cancer risk.
- 3 Confounder:** Many coffee drinkers are also **smokers**. Smoking, not coffee, is the real cause.



Summary of Epidemiology

- 1 Epidemiology is the study of the **distribution** and **determinants** of health-related outcomes in populations.
- 2 **Study design** is a key component of epidemiology (cohort, case-control).
- 3 **Relative risks and odds ratios** are used to measure association.
- 4 It is important to consider and address **bias** in epi studies.
- 5 Understanding **confounding** is important when conducting association studies.



Genetic Epidemiology

Genetic epidemiology is the study of the role of **genetic factors** in determining health and disease in **families and in populations**, and the interplay of such genetic factors with **environmental factors**.



Familial & Genetic Determinants

Genetic epidemiology focuses on **familial**, and in particular **genetic**, determinant of disease and the **joint effect of gene and environmental determinant**.

Monogenic

Single gene disorder or disease.

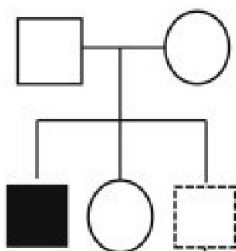
e.g., sickle cell anemia

Complex disease

Interacting between **genes and environment**.

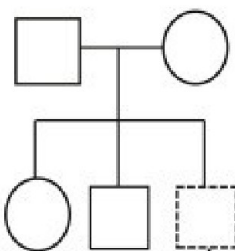
e.g., diabetes mellitus, ischemic heart disease, cancer

FAMILY AGGREGATION STUDIES



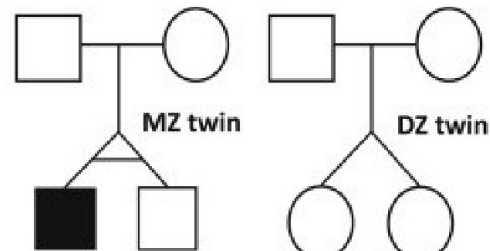
- Familial recurrence risk
- Relative recurrence risk ratio (λ)

ADOPTION STUDIES



- Phenotypic correlations

TWIN STUDIES



- Intraclass correlations in MZ vs DZ twins
- Concordance rates in MZ vs DZ twins
- Study of discordant phenotypes in MZ twins
- Cotwin-control studies



Why Study How Our Genome Is Involved in Disease?

1

Public Health

Identify people at **high risk** — people with a high risk of developing disease

2

Understanding Biology

Understand **disease pathways** — the association and causes of disease

3

Clinical Applications

Develop **better treatments** — develop new treatments

“Association does not imply causation”



Public Health: Identify People at High Risk

Identify people with a **high risk of developing disease.**

IN THE NEWS — MARCH 6, 2018

23andMe granted first FDA authorization for a direct-to-consumer genetic test on **cancer risk.**

CONDITIONS REPORTED

Parkinson's disease

a nervous system disorder impacting movement

Late-onset Alzheimer's disease

a progressive brain disorder that destroys memory and thinking skills

Celiac disease

a disorder resulting in the inability to digest gluten

Alpha-1 antitrypsin deficiency

a disorder that raises the risk of lung and liver disease

Early-onset primary dystonia

a movement disorder involving involuntary muscle contractions and other uncontrolled movements

Factor XI deficiency

a blood clotting disorder

Gaucher disease type 1

an organ and tissue disorder

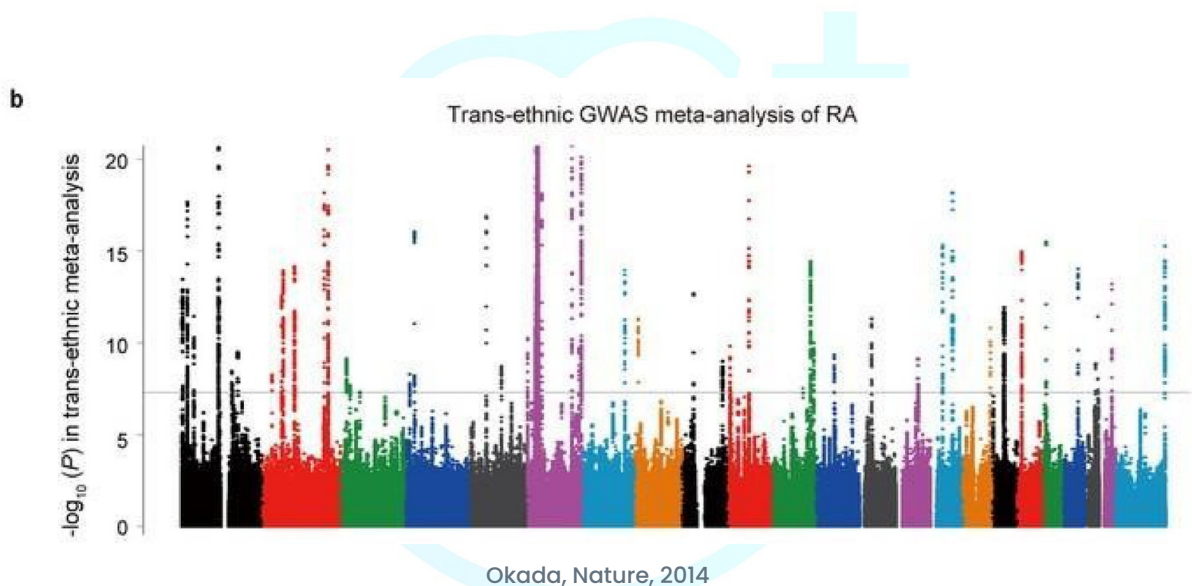
Glucose-6-Phosphate Dehydrogenase deficiency



Clinical Applications: Rheumatoid Arthritis

Rheumatoid Arthritis — an **inflammatory, crippling, incurable** disease. In 2005, an estimated **1.5 million (0.6%)** of US adults age ≥ 18 had RA.

A study of **10 million genetic variants** in **29,880 RA cases** and **73,758 controls**.



Identified genes are targets of **approved therapies for RA**, and further suggest that drugs approved for other diseases may be **repurposed (reuse)** for the treatment of RA.

- 1 **Flavopiridol** (originally for leukemia) → may also help in **liver cancer**.
- 2 **Capridine- β** (developed for cancer) → could be useful in **psoriasis/RA**.
- 3 **Ibrance (palbociclib)** (for breast cancer) → may also help in **other cancers or RA**.



Genetic Research Paradigm

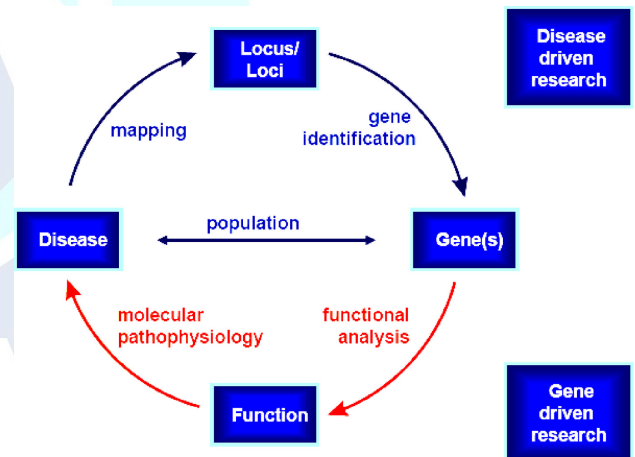
There are **two research approaches** in genetic epidemiology:

1. Disease-driven (blue cycle)

Start with the **disease** → map chromosomal regions (loci) → identify the gene(s) → study how it affects disease.

2. Gene-driven (red cycle)

Start with a **gene** → study its function (functional analysis) → understand how it leads to disease (molecular pathophysiology).



Both cycles connect through **population studies**, where genes and diseases are studied together.

We can approach genetic epidemiology from the **disease side** (find the gene) or from the **gene side** (find what disease it causes).



Study Designs in Genetic Epidemiology

The samples needed for genetic epidemiology studies may be **family-based** or **population-based**.



Family-based studies

- **Nuclear families** — index case and parents
- **Affected relative pairs** — sibs, cousins, any two members of the family
- **Extended pedigrees**
- **Twins** — compare identical vs fraternal twins (monozygotic and dizygotic)

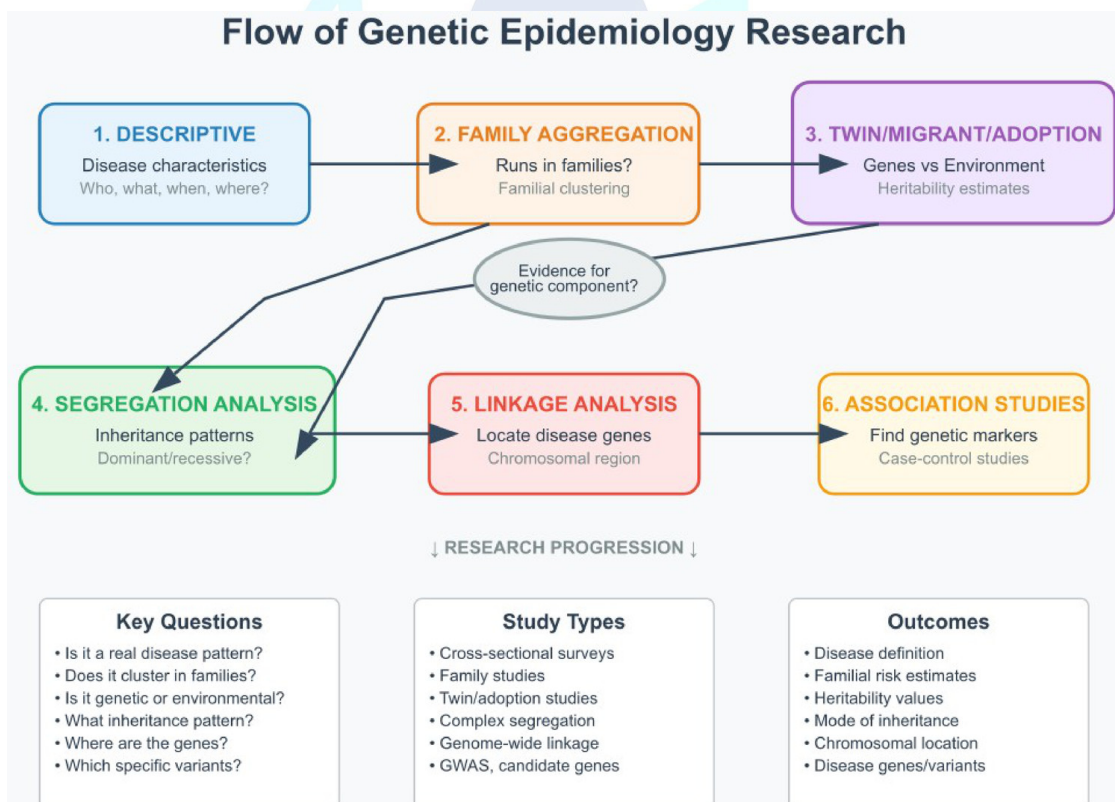
Population-based studies

Large samples of **unrelated individuals**.



Flow of Genetic Epidemiology Research

- 1 **Descriptive** — disease characteristics
- 2 **Family aggregation** — runs in families?
- 3 **Twin / migrant / adoption** — genes vs environment
- 4 **Segregation analysis** — inheritance patterns
- 5 **Linkage analysis** — locate disease genes
- 6 **Association studies** — find genetic markers



<http://www.dorak.info/epi/genetepi.html>

REMEMBER — AIM OF GENETIC EPIDEMIOLOGY

To detect the **inheritance pattern** of a particular disease, to **localize the gene** and to find a **marker associated with disease susceptibility**.



To Summarise...

DESIGN

2

Core study designs

- Cohort follows healthy people forward in time
- Case-control compares cases with controls
- Choice depends on how rare the disease is

MEASURE

RR & OR

Measuring association

- Relative Risk: exposed vs unexposed risk
- Odds Ratio: odds of exposure, cases vs controls
- RR for cohort studies, OR for case-control

THREATS

2

Bias and confounding

- Bias is systematic error in one direction
- Selection bias and recall bias are the two types
- A confounder distorts the true relationship

DISEASE

2

Types of genetic disease

- Monogenic: single gene, e.g. sickle cell anemia
- Complex: genes interacting with environment
- Diabetes, ischemic heart disease, cancer

WHY

3

Reasons to study the genome

- Public health: identify people at high risk
- Understanding biology: disease pathways
- Clinical: develop new and better treatments

FLOW

6

Research progression

- Descriptive, then family aggregation
- Twin/migrant/adoption, then segregation
- Linkage analysis, then association studies

THE AIM, IN ONE LINE

Detect the **inheritance pattern** of a disease, localize the gene, and find a marker associated with **disease susceptibility**.