Introduction to Epidemiology and Genetic Epidemiology
ep·i·de·mi·ol·o·gy
/ˌɛpəˌdēmēˈələjē/
noun
the branch of medicine that deals with the incidence, distribution, and possible control of diseases and other factors relating to health.
Major goals in Epidemiology

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

• To use this knowledge to prevent and treat disease
Key concepts in Epidemiology

• Incidence
  • Number of new cases in a population during a fixed time period
    • The reported number of new prostate cancer cases in United States during 2015 was 183,529.

• Prevalence
  • Number of existing cases in a population at a given time
    • In 2015, there were an estimated 3,120,176 men living with prostate cancer in the United States.
Cohort vs. case-control studies
Estimated incidence rates in cohorts

<table>
<thead>
<tr>
<th>Disease incidence per 100,000 per year (%)</th>
<th>Disease examples</th>
<th>Number of incident cases in 5 years for different cohort sizes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>200,000</td>
</tr>
<tr>
<td>10 (0.01)</td>
<td>Parkinson disease, schizophrenia</td>
<td>91</td>
</tr>
<tr>
<td>50 (0.05)</td>
<td>Colorectal cancer, renal failure</td>
<td>456</td>
</tr>
<tr>
<td>100 (0.10)</td>
<td>Breast cancer, hip fracture</td>
<td>912</td>
</tr>
<tr>
<td>200 (0.20)</td>
<td>Diabetes, stroke, heart failure</td>
<td>1,820</td>
</tr>
<tr>
<td>500 (0.50)</td>
<td>Myocardial infarction, all cancers</td>
<td>4,524</td>
</tr>
<tr>
<td>3,000 (3.00)</td>
<td>Cataracts, hypertension</td>
<td>25,858</td>
</tr>
</tbody>
</table>

Estimated numbers of incident cases available after 5 years of follow-up across the entire age range in the US population are shown, assuming an attrition rate of 3% per year. Data are taken from the Incidence and Prevalence Database.

Compared to cohorts, case-control studies are cheap, fast and powerful.

However, case-control studies suffer from several drawbacks:
- the need to identify appropriate controls
- they are more sensitive to recall bias
- it is difficult to assess rare exposures due to small sample sizes
Main Measures of Association in Epidemiology

• Relative Risk
  measure of the *relative probability of developing disease* given exposure status

• Odds Ratio
  measure of the *relative odds of exposure* given disease status (can approximate the Relative Risk when disease is rare)
<table>
<thead>
<tr>
<th>Exposure</th>
<th>Disease status</th>
<th>Cases</th>
<th>Controls</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed</td>
<td>a</td>
<td>b</td>
<td>a+b</td>
<td></td>
</tr>
<tr>
<td>Not Exposed</td>
<td>c</td>
<td>d</td>
<td>c+d</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>a+c</td>
<td>b+d</td>
<td>a+b+c+d</td>
<td></td>
</tr>
</tbody>
</table>
Relative Risk (RR) For Count Data

- Relative probability of developing disease given exposure status
- Used in cohorts
- Also known as risk ratio
- If no association RR=1

<table>
<thead>
<tr>
<th></th>
<th>Cases</th>
<th>Controls</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed</td>
<td>a</td>
<td>b</td>
<td>a+b</td>
</tr>
<tr>
<td>Not Exposed</td>
<td>c</td>
<td>d</td>
<td>c+d</td>
</tr>
</tbody>
</table>

\[
RR = \frac{\frac{a}{a+b}}{\frac{c}{c+d}} = \frac{(Incidence \ of \ Disease \ in \ Exposed)}{(Incidence \ of \ Disease \ in \ Unexposed)}
\]
Odds Ratio (OR) For Count Data

- Relative odds of exposure given disease status
- Used primarily in case-control studies
- Good estimate of RR
- If no association OR=1

\[
\begin{align*}
\text{OR} &= \frac{a/c}{b/d} = \frac{a*d}{b*c} = \frac{(Odds of Exposure among Cases)}{(Odds of Exposure among Controls)}
\end{align*}
\]
Confidence Intervals and p-values

• Relative risks and odds ratios give information on the magnitude of association

• Important to consider precision and statistical significance, along with estimate of magnitude of association.

• Statistical software will in addition to relative risks and odds ratios provide estimate of confidence intervals and p-values
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.
Sources of Bias in Epidemiology

Bias = Systematic error in the design, conduct or analysis of a study that results in a mistaken estimate of an exposure’s true effect on the risk of disease
Exercise and cardiovascular disease

Individuals who do not exercise tend to smoke more, have a more unhealthy diet and are more likely to have diabetes.

Your cases and controls might come from different underlying populations (e.g. men vs. women, old vs. young).

Individuals might misremember or give false information about their exercise routine.
Some common sources of bias

- **Selection Bias**
  - Arises when cases and controls are coming from different source populations (e.g. female cases, male controls)

- **Survival bias**
  - When cases are recruited some time after they were diagnosed. Might lead to a milder form of disease. This is especially true for aggressive/fatal disease (e.g. pancreatic cancer, heart attack)

- **Diagnostic bias**
  - If the investigator determining the outcome knows whether the person was exposed or not to the risk factor under study (e.g. if the radiologist knows that a potential pulmonary disease patient smokes, she may look more carefully at the x-ray).

- **Recall bias**
  - Accuracy and completeness of exposures, life style behaviors etc (e.g. cases might be more motivated to complete a questionnaire accurately).
Confounding

• A confounder is often defined as a factor that is:
  ① A risk factor for disease
  ② Associated with exposure
  ③ Not a direct result of exposure

• Confounding can lead to false positive findings.
Confounding example: Birth order and Down syndrome

Data from Stark and Mantel (1986)  
Source: Rothman 2002
Confounding example: Birth order and Down syndrome

• Later order children have higher risk
  • Maternal age is associated with birth order
  • Maternal age is associated with Down Syndrome

• Stratifying on maternal age, there is no longer evidence of an association between birth order and Down syndrome
Summary

• Epidemiology is the study of the distribution and determinants of health-related outcomes in populations

• Study design is a key component of epidemiology

• Relative risks and odds ratios are used to measure association

• It is important to consider and address bias in epi studies

• 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.
A gene that contributes to obesity has been identified for the first time, promising to explain why some people easily put on weight while others with similar lifestyles stay slim.

**Newsroom**

**Brain-Aging Gene Discovered**

Genetic variant accelerates normal brain aging in older people by up to 12 years

March 15, 2017

**HUTCH NEWS**

**Does aspirin prevent colorectal cancer? Depends on your DNA**

Fred Hutch researchers move closer to cracking the code on how genes and environmental factors influence colorectal cancer risk

March 17, 2015 | By Diane Mapas / Fred Hutch News Service

**Smoking addiction gene found**

Scientists say a gene makes people more likely to get hooked on tobacco, causing them to smoke more, making it harder to quit, and leading more often to deadly lung cancer. [Full story](http://www.newsweek.com/differing-conclusions)

Researchers make human-cow embryos

Science wishy-washy on water benefits | Vote
Why do we want to study how our genome is involved in disease?
Why do we want to study how our genome is involved in disease?

Public Health

Understanding Biology

Clinical Applications
Why do we want to study how our genome is involved in disease?

Public Health

Understanding Biology

Identify people with high risk of developing disease

Clinical Applications
My mother fought cancer for almost a decade and died at 56. She held out long enough to meet the first of her grandchildren and to hold them in her arms. But my other children will never have the chance to know her and experience how loving and gracious she was.

We often speak of “Mommy’s mommy,” and I find myself trying to explain the illness that took her away from us. They have asked if the same could happen to me. I have always told them not to worry, but the truth is I carry a “faulty” gene, BRCA1, which sharply increases my risk of developing breast cancer and ovarian cancer.
Specific inherited mutations in the BRCA1 and BRCA2 genes increase the risk of breast and ovarian cancers. Testing for these mutations is usually recommended in women without breast cancer only when the person's individual or family history suggests the possible presence of a harmful mutation in BRCA1 or BRCA2. Testing is often recommended in younger women newly diagnosed with breast cancer because it can influence treatment decisions and have implications for their family members.
## Health Risks (122)

<table>
<thead>
<tr>
<th>Condition</th>
<th>Your Risk</th>
<th>Average Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coronary Heart Disease</td>
<td>33.1%</td>
<td>24.4%</td>
</tr>
<tr>
<td>Psoriasis</td>
<td>15.0%</td>
<td>10.1%</td>
</tr>
<tr>
<td>Restless Legs Syndrome</td>
<td>5.2%</td>
<td>4.2%</td>
</tr>
<tr>
<td>Exfoliation Glaucoma</td>
<td>2.9%</td>
<td>1.0%</td>
</tr>
<tr>
<td>Lupus (Systemic Lupus Erythematosus)</td>
<td>1.1%</td>
<td>0.2%</td>
</tr>
</tbody>
</table>

See all 122 risk reports...

## Inherited Conditions (53)

<table>
<thead>
<tr>
<th>Condition</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hemochromatosis (HFE-related)</td>
<td>Variant Present</td>
</tr>
<tr>
<td>ARSACS</td>
<td>Variant Absent</td>
</tr>
<tr>
<td>Agenesis of the Corpus Callosum with Peripheral Neuropathy (ACCPN)</td>
<td>Variant Absent</td>
</tr>
<tr>
<td>Alpha-1 Antitrypsin Deficiency</td>
<td>Variant Absent</td>
</tr>
<tr>
<td>Autosomal Recessive Polycystic Kidney Disease</td>
<td>Variant Absent</td>
</tr>
</tbody>
</table>

See all 53 carrier status...

## Traits (62)

<table>
<thead>
<tr>
<th>Trait</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alcohol Flush Reaction</td>
<td>Does Not Flush</td>
</tr>
<tr>
<td>Bitter Taste Perception</td>
<td>Can Taste</td>
</tr>
<tr>
<td>Blond Hair</td>
<td>28% Chance</td>
</tr>
<tr>
<td>Earwax Type</td>
<td>Wet</td>
</tr>
<tr>
<td>Eye Color</td>
<td>Likely Blue</td>
</tr>
</tbody>
</table>

See all 62 traits...

## Drug Response (25)

<table>
<thead>
<tr>
<th>Drug Name</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clopidogrel (Plavix®) Efficacy (CYP2C19-related)</td>
<td>Reduced</td>
</tr>
<tr>
<td>Abacavir Hypersensitivity</td>
<td>Typical</td>
</tr>
<tr>
<td>Acetaldehyde Toxicity</td>
<td>Typical</td>
</tr>
<tr>
<td>Fluorouracil Toxicity</td>
<td>Typical</td>
</tr>
<tr>
<td>Hepatitis C Treatment Response</td>
<td>Typical</td>
</tr>
</tbody>
</table>

See all 25 drug response...
• **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**, also known as G6PD, a red blood cell condition;
• **Hereditary hemochromatosis**, an iron overload disorder; and
• **Hereditary thrombophilia**, a blood clot disorder.
Why do we want to study how our genome is involved in disease?

- Understanding Biology
  - Understand the causes of disease
- Public Health
- Clinical Applications
“Association does not imply causation”
HDL ("Good") Cholesterol and Myocardial Infarction (MI)

• ↑ HDL -> ↓ MI risk

Increasing HDL concentrations might help decrease cardiovascular disease risk.
People who carry gene variants that increase HDL do not have a lower risk of MI since HDL is correlated with exercise, weight loss, diet (nuts, fish) it is likely that these lower your risk for MI rather than HDL itself.
Rare variant in scavenger receptor BI raises HDL cholesterol and increases risk of coronary heart disease
Why do we want to study how our genome is involved in disease?

- Understanding Biology
- Public Health
- Develop new treatments
- Clinical Applications
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 for the treatment of RA.

Lymphoma/Leukemia/Liver cancer

Cancers/psoriasis

flavopiridol

Breast cancer

flavopiridol

Cancers/psoriasis

<table>
<thead>
<tr>
<th>Cancer Type</th>
<th>Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRC</td>
<td>58, 79, 52</td>
</tr>
<tr>
<td>BC</td>
<td>70</td>
</tr>
<tr>
<td>Stroke</td>
<td>81</td>
</tr>
<tr>
<td>CHD</td>
<td>63</td>
</tr>
</tbody>
</table>

**Key**

- **Male**: □
- **Affected male**: ■
- **Deceased male**: ×
- **Female**: ○
- **Affected Female**: ●
- **Deceased Female**: Ø
- **Affected female with different cancer**: ☿