Epidemiology for the Boards

Khalil Ghanem, MD, PhD
Professor
Johns Hopkins University School of Medicine
kghanem@jhmi.edu

1

Disclosures

- None

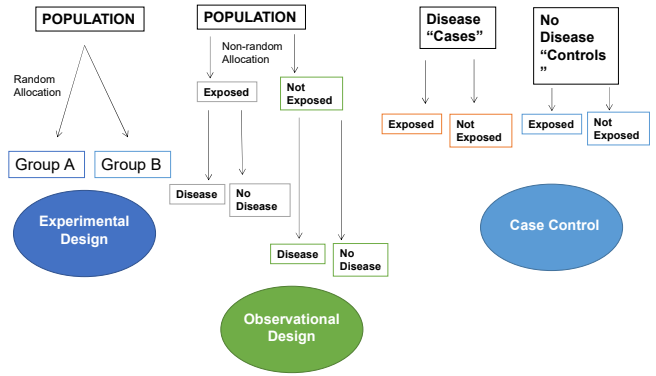
2

Overview

- Study designs
- Incidence & Prevalence
- Relative risk, relative odd, & attributable risk
- Confidence intervals
- Number needed to treat
- Sensitivity, specificity, positive predictive value, negative predictive value
- Bias and confounding

3

Study Designs



4

©2026 Infectious Disease Board Review, LLC

Example: Study Designs

- **Choose the most appropriate study design for the following scenarios:**
  - You are trying to determine what caused 35 people to experience fever and severe hemorrhagic complications upon returning from a Caribbean cruise
  - You want to get FDA approval for a novel influenza vaccine
  - You want to determine whether hormonal contraception increases your risk of HIV

5

Incidence vs. Prevalence

- **Incidence** = *new* infection occurring during a specified period of time in a population at risk for developing the infection
  - A measure of events (a disease that develops in someone who did not have it), thus, a measure of *risk*
- **Prevalence:** number of affected persons present in the population at a given time(i.e. *existing* infections)
- **Prevalence=Incidence X duration of disease**

6

Example: Incidence vs. Prevalence

- **In a population that includes persons with HIV who exhibit high medication adherence, what would the impact of ART be on HIV incidence and prevalence over a 10-year period?**
  - Incidence = new HIV infections. ART should decrease the risk of transmission of HIV and thereby **decrease** the incidence
  - Prevalence = all existing HIV infections. ART allows people with HIV to live longer so it may **increase** the prevalence of HIV

7

Estimating Risk

- **Relative Risk (RR)**= Incidence in exposed/ Incidence in nonexposed
  - If the RR=1, there is no association
  - If the RR >1, the risk in exposed > nonexposed
  - If the RR<1, the risk in exposed < nonexposed
- **Hazards Ratio(HR)**: A form of RR; HR is instantaneous while RR is cumulative.
- **Odds**= Probability that disease developed/Probability that it did not develop
- **Odds Ratio:**
  - **Cohort study:** ratio of odds of disease occurring in exposed to the odds of disease occurring in non-exposed
  - **Case Control**: ratio of the odds that the cases were exposed to the odds that the controls were exposed
  - If the OR=1, there is no association between exposure and disease
  - If the OR>1, the exposure is positively related to the disease
  - If the OR<1, the exposure is negatively related to the disease

8

©2026 Infectious Disease Board Review, LLC

Example: Estimating Risk

- **In a population of 1000 people, 400 were having condomless sex. Infection-Y occurred in 100 of the 400 who were having condomless sex and in 5 of the 600 who were not.**
- **What is the RR of Y in those having condomless sex?**
- **What are the relative odds (odds ratio) of Y in those having condomless sex?**

- RR: 100/400/5/600= 31.3
- OR: 100/300/5/595=41.3

- The odds ratio is a good estimate of the relative risk when the disease being studied is RARE

9

Estimating Risk 2

- The **attributable risk** is the proportion of disease incidence that can be attributed to a specific exposure

  AR= Incidence in exposed- Incidence in non-exposed

- This is one of the most important measures when deciding *how* to spend money and resources in public health

10

Example: Estimating Risk 2

A new deadly fungal infection is described with a mortality rate of 30%.

You are given 1 million dollars to spend on prevention in your state.

- Persons with Exposure A have a RR of 16 for getting infected.
- Persons with Exposure B have a RR of 2 for getting infected.

How will you spend your money?

11

Example: Estimating Risk 2

- Exposure A is spelunking and Exposure B is gardening
  - **NOW how are you going to spend your money?**
- Even though the relative risk of spelunking is far more than gardening, most of the cases in your state are likely the result of gardening (a lot more people garden).
- The attributable risk of gardening, therefore, is much greater than that of spelunking

Exposure	Incidence	Relative Risk	Attributable Risk
Spelunking	32 per million	**16**	30 per million
No Spelunking	2 per million		
Gardening	640 per million	2	**320 per million**
No Gardening	320 per million		

12

©2026 Infectious Disease Board Review, LLC

Confidence Intervals

- Confidence intervals (CI) are used to indicate the reliability of an estimate
  - CI is *directly* related to the standard deviation and *indirectly* related to the sample size (i.e., the larger the sample size, the smaller the CI)
- In simple terms, a 95% CI means: If you were to repeat this experiment many times, 95% of the time, your results will fall within this range.
  - The wider the CI surrounding the point estimate, the more uncertainty there is about the reliability of that point estimate

13

Example: Confidence Intervals

- Match each scenario to the more likely prevalence point estimate and CI:
  - **Scenario 1**: We test 100 people in the population for HIV.
  - **Scenario 2**: We test 3500 people in the population for HIV

A. The prevalence of HIV is 1.3% (95%CI: 1.1 %-1.5%)
B. The prevalence of HIV is 3.3% (95%CI: 0.3%-7.2%)

14

Number Needed to Treat (NNT)

- NNT= 1/(Rate in untreated)- (Rate in treated)

15

Example: NNT

RCT for a new Ebola vaccine: the mortality rate in the experimental group is 20 per 100 while the mortality rate in the control group is 85 per 100. How many people do we need to vaccinate to prevent one death from Ebola?

NNT= 1/(0.85-0.20)=1.5

1.5 people need to be vaccinated to prevent a single death from Ebola. This would be a GREAT public health intervention in endemic areas.

16

©2026 Infectious Disease Board Review, LLC

Sensitivity, Specificity, Positive Predictive Value (PPV), Negative Predictive Value (NPV)

	Disease	No Disease
Positive	True Positive	False positive
Negative	False negative	True negative

Sensitivity= TP/ TP + FN
Specificity= TN/ TN +FP
PPV= TP/ TP + FP
NPV= TN/TN +FN

Sensitivity and specificity are INDEPENDENT of prevalence whereas PPV and NPV are DEPENDENT on prevalence

- **Sensitivity** = the ability of a test to correctly identify those who have a disease
- **Specificity** = the ability of a test to correctly identify those who do not have a disease
- **PPV**= the proportion who test positive and actually have the disease
- **NPV**=the proportion who test negative and actually don't have the disease

17

Example: Sensitivity Specificity, PPV, NPV

The glycoprotein-G- based antibody tests for the detection of HSV-2 antibodies have a sensitivity of 99% and specificity of 98.5%. We plan to test two populations: (A) 1000 commercial sex workers (B) 1000 nuns confined to a convent.

In which population will the tests have a higher: Sensitivity? Specificity? PPV? NPV?

- Sensitivity and specificity are INDEPENDENT of prevalence of disease. As such, the sensitivity and specificity of these tests will be the same in both populations
- Population A likely has a higher prevalence of HSV-2 compared to population B. As such, the PPV of the test will be higher in population A and the NPV will be higher in population B

18

Definitions

- **Precision**: How close do the results cluster to ***each other***?
- **Accuracy**: How close do the results cluster to ***the truth***?
- **Bias**: systematic error leading to a decrease in accuracy
  - Bias is reduced by careful study design
- **Confounding**: a distortion in the degree of association between an exposure and an outcome due to a mixing of effects between the exposure and an incidental factor, which is known as the confounder
  - **You must adjust for confounding; otherwise, it will lead to misinterpretation of results**
- **Effect Modification** (i.e., interaction): a variable that differentially (positively and negatively) modifies the observed effect of a risk factor on disease status. Different groups have different risk estimates when effect modification is present
  - **Effect modification is a true phenomenon that should be reported. You do NOT need to adjust for it.**

19

Example: Definitions

- Drinking coffee is found to be strongly associated with an increased risk of HPV-induced cervical cancer. We later find out that those who drink coffee are much more likely to smoke cigarettes.
- Cigarette smoking is a _____ in the relationship between coffee drinking and cervical cancer

20

©2026 Infectious Disease Board Review, LLC

Thank you!

©2026 Infectious Disease Board Review, LLC