STAT931: Causal Inference and Epidemiological Studies
Lecture Notes

Yiliu Cao

Sept 13, 2024

Contents

1 Introduction

1.1 Quantities of Interest

In Epidemiological Studies, we want to measure the occurence. More specifically, we want to measure the **prevalence** and **incidence**.

Prevalence

There are two measures of prevalence

1. **Point Prevalence**: The proportion of at-risk population affected at a specific time **point**. Formally:

2. **Period Prevalence**: The proportion of at-risk population affected at a specific time **period**.

Incidence

There are also two measures of incidence.

1. **Incidence Proportion**: The proportion of a defined at-risk population who has affected within a specific time **point**.

2. **Incidence Rate**:

1.2 Association

Association is **NOT** causation, it is the comparisions between groups.

Definition 1.1. Relative Risk of an outcome Y with a binary risk factor A is:

$$RR = \frac{P(Y = 1 \mid A = 1)}{P(Y = 1 \mid A = 0)}$$

where $0 < RR < \infty$

Note that:

1. $P(A = 1) + P(A = 0) = 1$

2. $RR = 1 \implies Y = 1 \perp A = 1 \implies$ No association between Y and A

3. If $RR > 1$, there is greater risk of $Y = 1$ when $A = 1$ vs $A = 0, vice - versa$

4. $\frac{P(Y=1|A=1)}{P(Y=1|A=0)} \neq \frac{P(A=1|Y=1)}{P(A=1|Y=0)}$

Definition 1.2. Odds Ratio

- The odds of a disease Y among the exposed $A = 1$

$$= \frac{P(Y = 1 \mid A = 1}{P(Y = 0 \mid A = 1)}\Bigg) = \frac{P(Y = 1 \mid A = 1)}{1 - P(Y = 1 \mid A = 1)}$$

- The odds of disease Y among the unexposed $A = 0$

$$= \frac{P(Y=1 \mid A=0)}{P(Y=0 \mid A=0)} \frac{P(Y=1 \mid A=0)}{1-P(Y=1 \mid A=0)}$$

- The odds ratio for measuring the association of disease with the exposed vs. unexposed groups is:

$$\begin{aligned} OR &= \frac{P(Y=1 \mid A=1)/P(Y=0 \mid A=1)}{P(Y=1 \mid A=0)/P(Y=0 \mid A=0)} \\ &= \frac{P(Y=1 \mid A=1)/[1-P(Y=1 \mid A=1)]}{P(Y=1 \mid A=0)/[1-P(Y=1 \mid A=0)]} \\ &= \underbrace{\frac{P(Y=1 \mid A=1)}{P(Y=1 \mid A=0)}}_{RR} \times \frac{P(Y=0 \mid A=0)}{P(Y=0 \mid A=1)} \end{aligned}$$

Note that:

1. $OR = 1 \Leftrightarrow log(OR) = 0$ means no association between Y and A
2. $OR \leq 1$ means greater odds of ratio when exposed
3. The OR for Y given A is equal to the OR for A given Y
$$OR = \frac{\frac{P(Y=1|A=1)}{P(Y=0|A=1)}}{\frac{P(Y=1|A=0)}{P(Y=0|A=0)}} = \frac{\frac{P(A=1|Y=1)}{P(A=0|Y=1)}}{\frac{P(A=1|Y=0)}{P(A=0|Y=0)}}$$
4. OR is a goods estimate of RR for rare disease as the second term above will approximate to 0.
5. $RR > 1 \implies OR > RR$, $RR < 1 \implies OR < RR$

Definition 1.3. Risk difference

$$RD = P(Y=1 \mid A=1) - P(Y=1 \mid A=0)$$

where $-1 < RD < 1$

This can be understood as the additional risk when exposed. In addition, a positive RD means a greater risk when exposed, vice-versa.

Definition 1.4. Attributable Risk(AR)
AR is the fraction of the cases of the outcome $Y = 1$ that can be attributed to $A = 1$

$$AR = \frac{P(Y=1) - P(Y=1 \mid A=0)}{P(Y=1)}$$

Furthermore:

$$\begin{aligned} AR &= \frac{P(A=1)[P(Y=1 \mid A=1) - P(Y=1 \mid A=0)]}{[P(Y=1 \mid A=1)P(A=1)] + [P(Y=1 \mid A=0)P(A=0)]} \\ &= \frac{P(A=1)[RR-1]}{[P(A=1)RR] + P(A=0)} = \frac{P(A=1)[RR-1]}{(P(A=1)[RR-1]) + 1} \end{aligned}$$

We can see that the AR depends on both the association between A and Y via RR, and the prevalence of risk factor at $A = 1$

Note that:

1. $AR = 0 \Longleftrightarrow RD = 0 \Longleftrightarrow OR = 1 \Longleftrightarrow RR = 1$ is null, meaning that $A \perp Y$
2. $AR > 0$ means the A will rise the risk of Y; $AR < 0$ means A is protective, decrease the risk of Y.
3. AR does not imply causation.