HARVARD T.H. CHAN SCHOOL OF PUBLIC HEALTH

STANLEY CENTER FOR PSYCHIATRIC RESEARCH AT BROAD INSTITUTE

Module 1: Introduction and Measures in Epidemiologic Studies

Part 3: Measures of Disease Occurrence

Bizu Gelaye, PhD, MPH
Departments of Epidemiology and Psychiatry
Harvard T.H. Chan School of Public Health and Harvard Medical School

Module 1: Introduction and Measures in Epidemiologic Studies

1. What is Epidemiology?
2. Epidemiological Concepts
3. **Measures of Disease Occurrence**
4. Measures of Association
5. Measures of Impact

Part 3: Measures of Disease Occurrence

Learning objectives:

- Understand what measures of disease occurrence are;
- Learn how to calculate and interpret:
  - Prevalence
  - Cumulative incidence (CI) and incidence rate (IR)
  - Mortality and case fatality
  - Odds
- Understand key differences between measures of occurrence

Measures of Disease Occurrence

- **Measures of occurrence** describe the distribution of outcome in a group of people with regards to person, place, and time

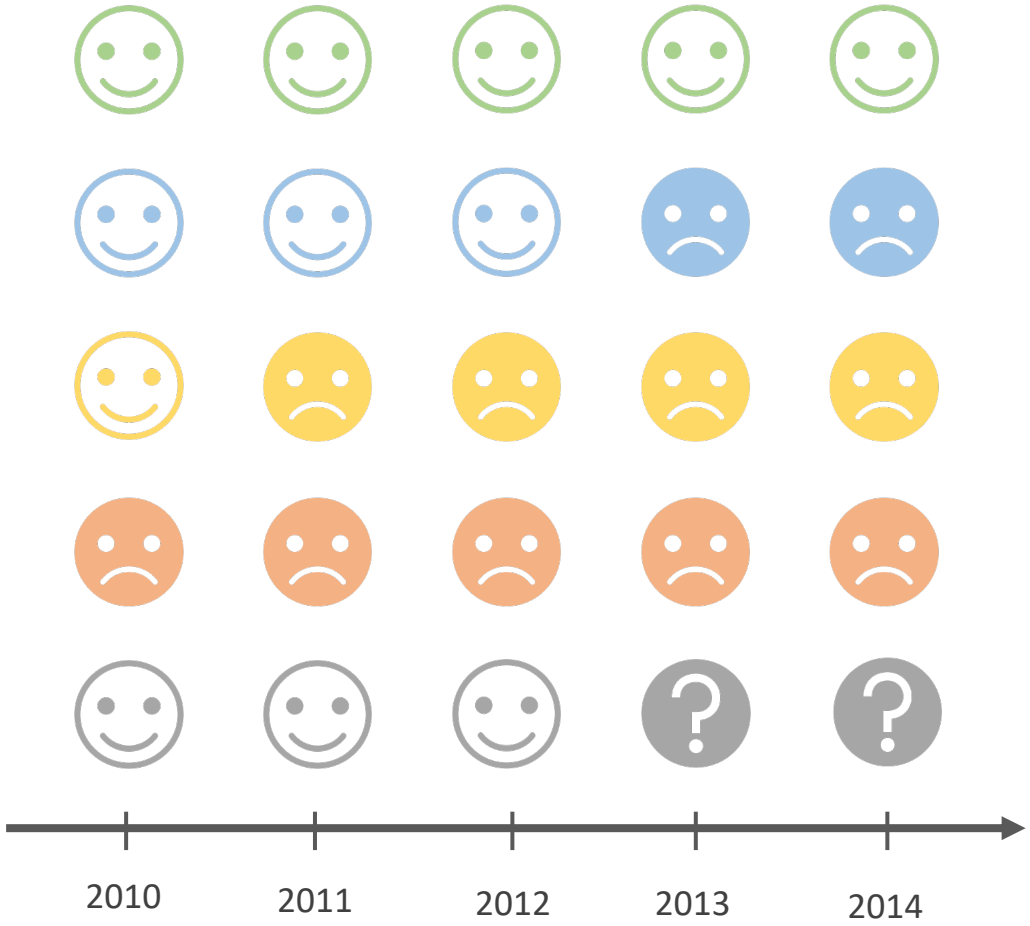
- They should take into account:
  - **Number** of people with the disease
  - **Size** of the population
  - **Length of time** people were followed

- As we discuss different measures of occurrence, keep these above in mind

Types of Variables

- **Continuous**
  - Examples: height, blood pressure, age
  - Distribution is usually described using mean, median, or percentiles
- **Discrete**
  - Examples: blood types, death, disease diagnosis
  - Distribution is usually described using proportion, odds, or rate
  - A discrete variable with only two possible values is often referred as a binary variable
- For our introduction to occurrence measures, we will be focusing on:
  - Binary exposure (yes, no)
  - Binary outcome (yes, no)

A hypothetical study population:



Outcome = No

Outcome = Yes

Outcome = unknown

Measures of Disease Occurrence

- Prevalence
- Incidence: cumulative incidence (CI) and incidence rate (IR)
- Mortality and case fatality
- Odds

Prevalence

- Proportion of population with existing disease/characteristic at a given point in time (or period) -- a “snapshot” of disease frequency

$$\text{Prevalence} = \frac{\#\ of\ existing\ cases}{\#\ of\ individuals\ in\ the\ study\ population} \text{ at a given point in time}$$

- Prevalence will be a number between 0 and 1 (i.e. a dimensionless proportion)
- Example:
  - The prevalence of dementia among individuals aged 71 and older was 13.9% in the USA in 2002

A hypothetical study population:

Prevalence in 2011 is:

$$\frac{2}{5} = 0.4$$

Prevalence during 2010-2014 is:

$$\frac{3}{5} = 0.6$$

Prevalence - Example

- A nationally representative household survey was conducted between 2002 and 2004 using the World Health Organization Composite International Diagnostic Interview (CIDI) to establish a diagnosis of depression. The dataset included 4,351 adult South Africans of all racial groups. It was found that 439 persons had lifetime depression. (Tomlinson, 2009)

- **What is the prevalence of persons having a major depressive episode in their lifetime?**

Prevalence - Example

- A nationally representative household survey was conducted between 2002 and 2004 using the World Health Organization Composite International Diagnostic Interview (CIDI) to establish a diagnosis of depression. The dataset included 4,351 adult South Africans of all racial groups. It was found that 439 persons had lifetime depression. (Tomlinson, 2009)

- **What is the prevalence of persons having a major depressive episode in their lifetime?**

$$\text{Prevalence} = \frac{439}{4351} = 0.1$$

Why prevalence?

Prevalence is useful when:

- Interested in existing disease at a certain point in time (quantify disease burden)
- May be easier to collect than other measures

Prevalence may <u>not</u> be useful when:

- Interested in the development or rate of a disease
- Interested in disease etiology: unclear temporality
- Prevalence is determined by occurrence and duration. High prevalence may be due to:
  - High occurrence, or
  - Long duration (survival)

- Next, we will discuss incidence measures that capture the development of the disease

Incidence

Two main types of incidence:

- Cumulative incidence (CI)
- Incidence rate (IR)

Cumulative Incidence (CI)

- Proportion of people with incident (new) events in a population over a specified time period

$$\text{Cumulative incidence} = \frac{\#\ of\ people\ who\ develop\ disease\ in\ a\ specified\ time\ period}{\#\ of\ individuals\ at\ risk\ of\ the\ disease\ at\ the\ start\ of\ the\ period}$$

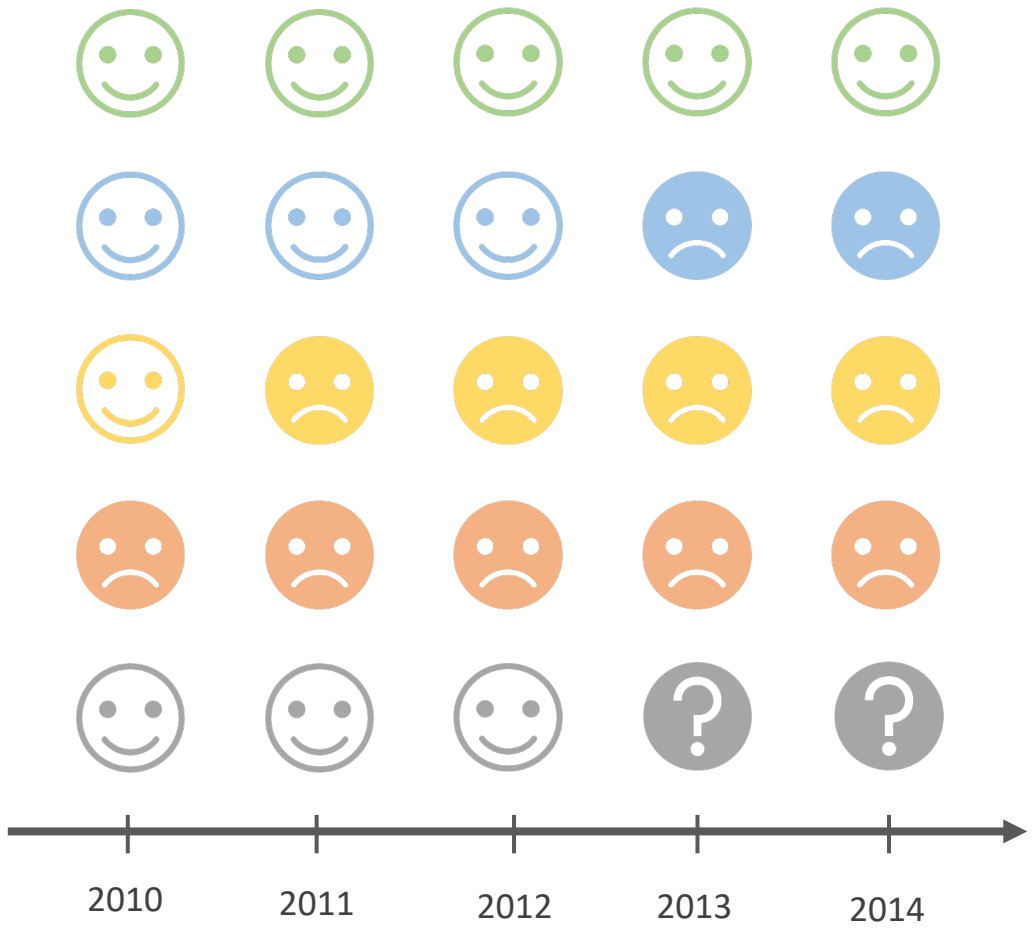
- Those in this population are considered “at risk” for the disease (i.e. disease free at start of follow-up)
- Follow the population over time and record if a new disease occurs
- Anyone included in the denominator must be eligible to move into the numerator

Cumulative Incidence (CI)

$$\text{Cumulative incidence} = \frac{\#\ of\ people\ who\ develop\ disease\ in\ a\ specified\ time\ period}{\#\ of\ individuals\ at\ risk\ of\ the\ disease\ at\ the\ start\ of\ the\ period}$$

- This is also a dimensionless proportion, so it will be between 0 and 1
- Must specify time period
  - CI increases over time, as cases of disease accumulate
  - E.g., the CI of death over 100 years is nearly 1!
- Assumes complete follow-up (i.e. no study dropout)
- Sometimes also referred as: incidence proportion, risk, attack rate

A hypothetical study population:



Cumulative incidence during 2010-2014 is

$$\frac{2}{5-1} = 0.5$$

Cumulative Incidence (CI) – Example 1

- In 1977, 750 entering students at National Taiwan University were found to be susceptible to hepatitis B (i.e., had no antibody, not a carrier)
- By 1981, 39 of them had seroconverted—i.e., had been infected with HBV since originally tested

- **What is the four-year cumulative incidence?**

Cumulative Incidence (CI) – Example 1

- In 1977, 750 entering students at National Taiwan University were found to be susceptible to hepatitis B (i.e., had no antibody, not a carrier)
- By 1981, 39 of them had seroconverted—i.e., had been infected with HBV since originally tested

- **What is the four-year cumulative incidence?**

$$\text{Cumulative incidence} = \frac{39}{750} = 0.052$$

Cumulative Incidence (CI) – Example 2

- A study recruits 1,000 women 65 and older for a study of Alzheimer's disease. Upon initial examination, the investigators discover that 15 of the potential study subjects already had Alzheimer's disease. This cohort is followed for seven years during which time 90 study subjects develop Alzheimer's disease.

- **What is the seven-year cumulative incidence of Alzheimer's disease?**

Cumulative Incidence (CI) – Example 2

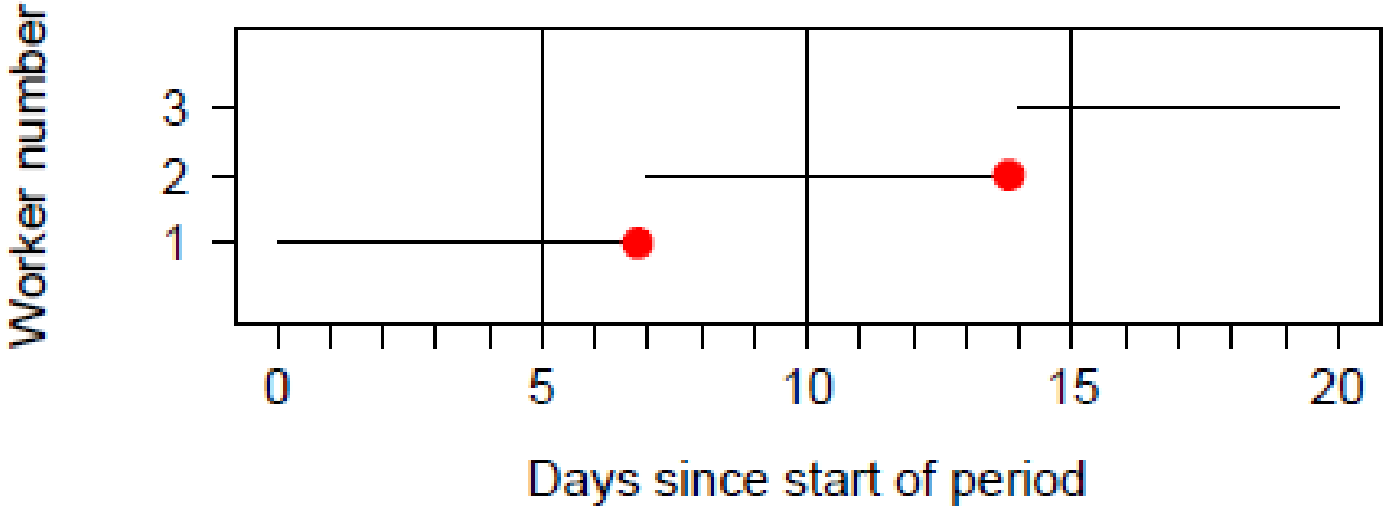
- A study recruits 1,000 women 65 and older for a study of Alzheimer's disease. Upon initial examination, the investigators discover that 15 of the potential study subjects already had Alzheimer's disease. This cohort is followed for seven years during which time 90 study subjects develop Alzheimer's disease.

- **What is the seven-year cumulative incidence of Alzheimer's disease?**

$$\text{Cumulative incidence} = \frac{90}{1000 - 15} = 0.09$$

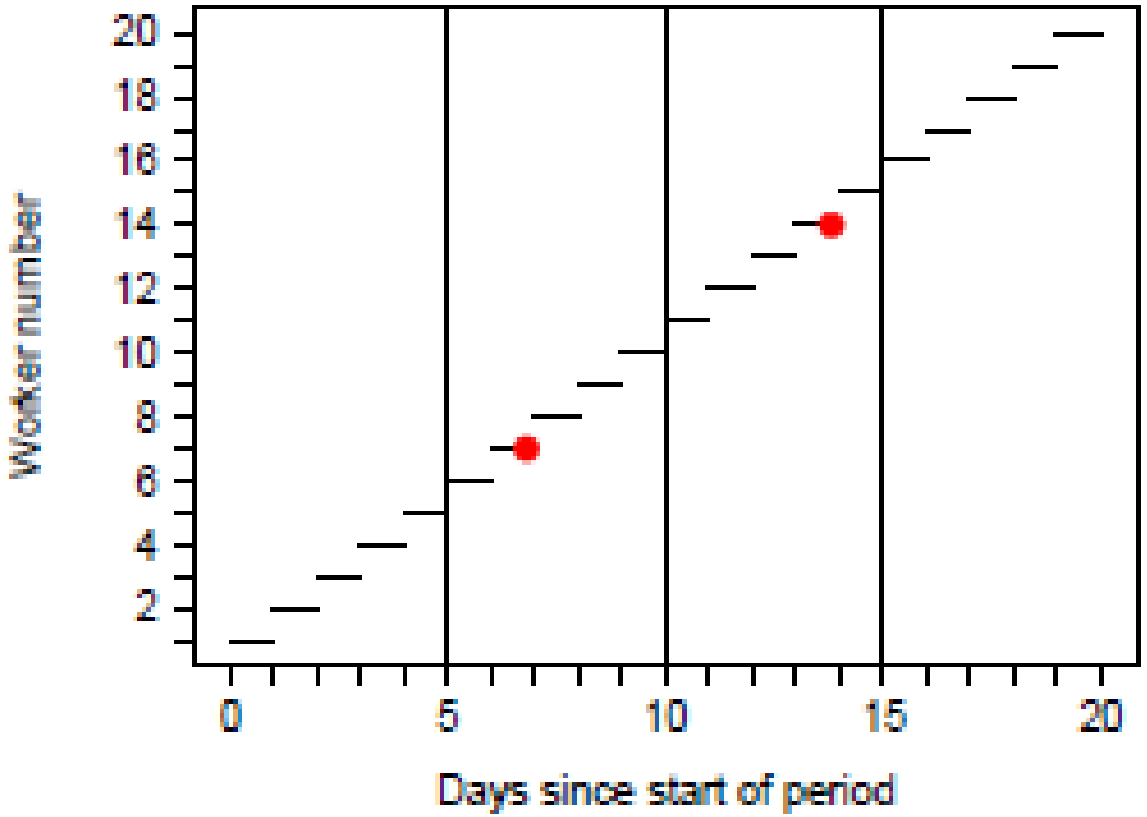
Incidence Rate (IR) – Motivating Example

- **Consider injuries among factory workers who operate a dangerous machine**
- **First 4-week period**
  - Worker 1 operates machine until injured on day 7
  - Worker 2 takes over until injured on day 14
  - Worker 3 takes over through day 20



Incidence Rate (IR) – Motivating Example

- **Second 4-week period**
  - Each day, a different worker operates the machine
  - Injuries occur on day 7 and 14



Incidence Rate (IR) – Motivating Example

- **The factory owner claimed that:** “Great progress on safety: an operator’s risk of injury was reduced more than 6-fold!”

Period	# of operators	# of injuries	Percent injured
First	3	2	67%
Second	20	2	10%

- **Do you agree? Is cumulative incidence an appropriate measure to use in this example? Why or why not?**

Incidence Rate (IR)

- Measures how quickly people are developing the disease at any point in time

$$\text{Incidence rate} = \frac{\#\ of\ people\ who\ develop\ disease}{\#\ of\ person-time\ when\ people\ were\ at\ risk\ of\ getting\ disease}$$

- Again, we focus on an “at risk” population (disease-free)
- Here, we follow the population of interest over time and record when a new disease event occurs

Person-time of Observation

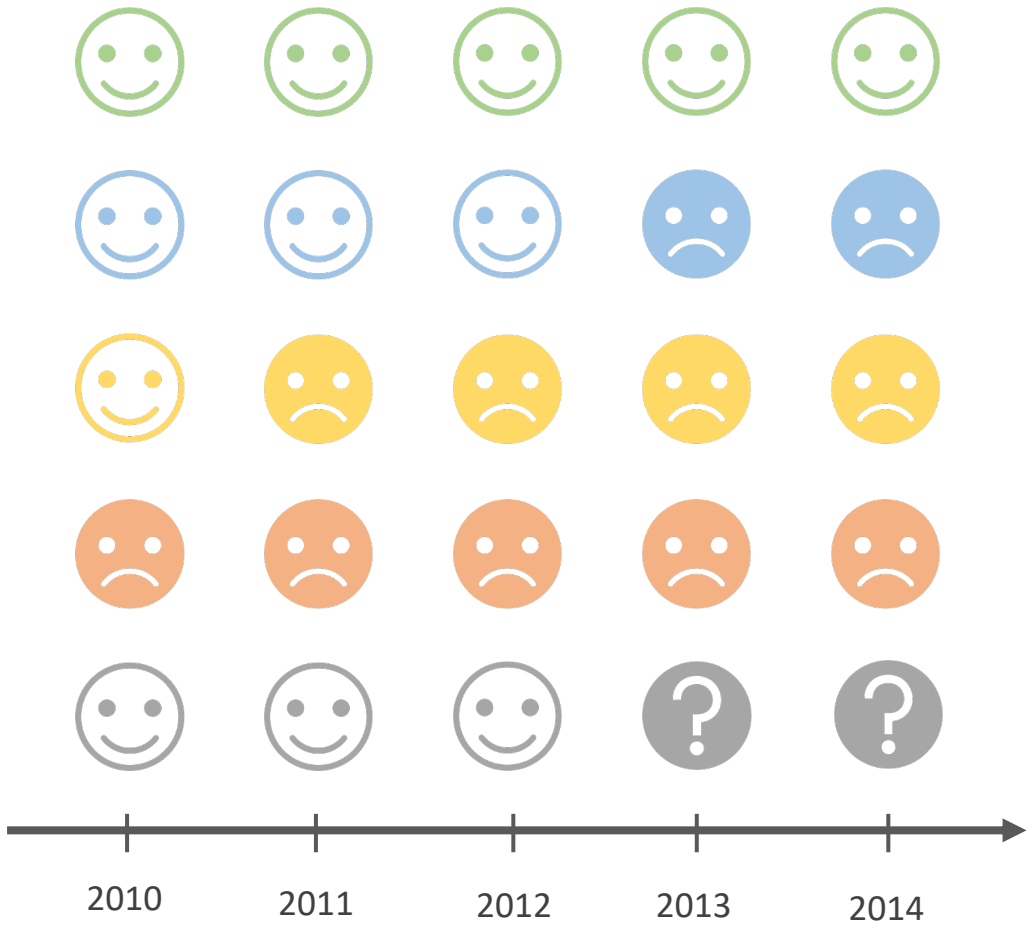
- Person-time is the sum of people's time at risk during a study period
- For each individual, the time at risk spans from the start of follow up to:
  - Outcome event occurs, or
  - Dropping out of study, or
  - Competing events (e.g. death), or
  - End of study
- For example, 50 person-years can be contributed by:
  - 50 persons followed for 1 year, or
  - 10 persons followed for 5 years, or
  - 1 person followed for 50 years

Incidence Rate (IR)

$$\text{Incidence rate} = \frac{\#\ of\ people\ who\ develop\ disease}{\#\ of\ person-time\ when\ people\ were\ at\ risk\ of\ getting\ disease}$$

- This is NOT a proportion; it can be greater than 1 (ranging from 0 to infinity).
- It must be interpreted with a time unit:
  - 1.5 cases/person-year = 15 case/10 person-years
  - 1 case/person-month = 12 cases/person-year
- Does not assume complete follow-up
- Can be calculated in closed cohorts and open cohorts (more in the next module)
- Sometimes also referred as: incidence density

A hypothetical study population:



Incidence rate during 2010-2014 is

$$\frac{2}{5+3+1+3} = 0.17 \text{ (or 17 per 100 person-years)}$$

The cumulative incidence during 2010-2014 is:

$$\frac{2}{5-1} = 0.5$$

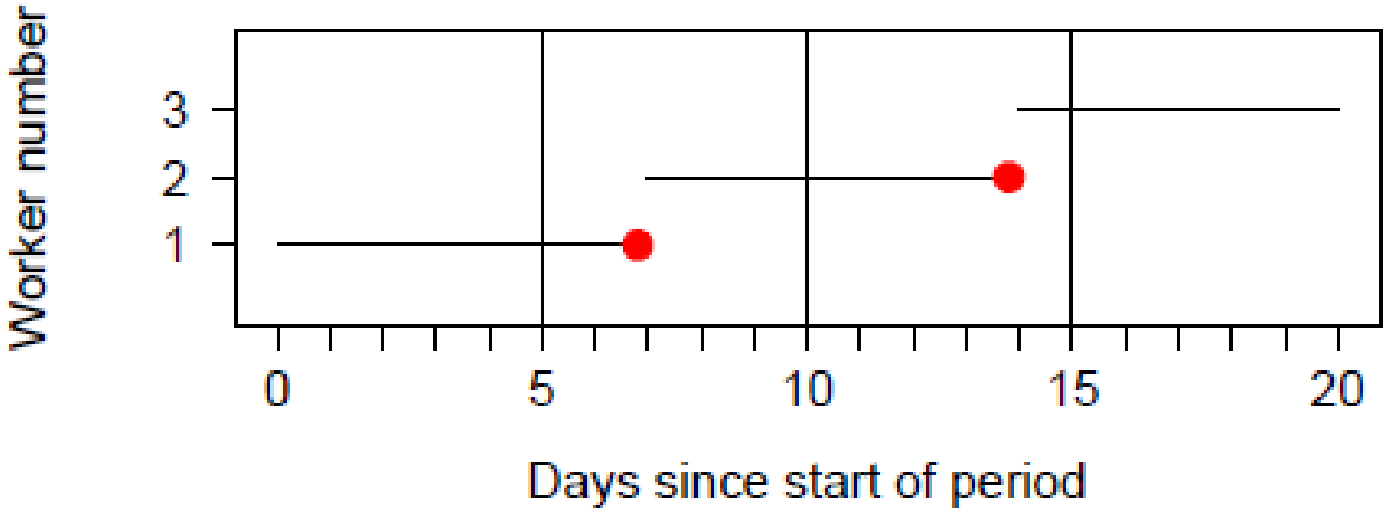
IR tells us how quickly people are developing the disease; however, CI is not sensitive to the timing of disease occurrence.

Going back to the motivating example

- **First 4-week period**

$$\text{IR} = \frac{No.of\ incident\ cases}{Amout\ of\ person{-}time\ at\ risk}$$

$$\text{IR} = \frac{1+1}{7+7+6}$$

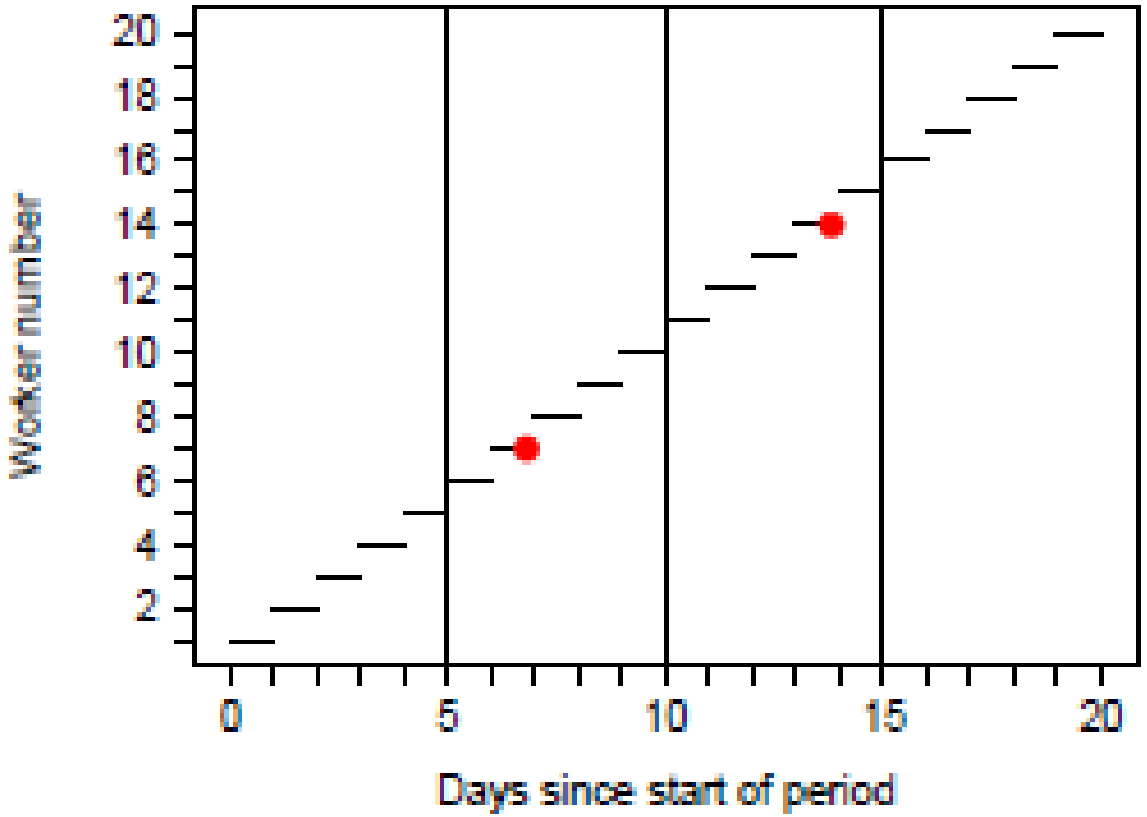


Incidence Rate (IR) – Motivating Example

- **Second 4-week period**

$$\text{IR} = \frac{No.of\ incident\ cases}{Amout\ of\ person{-}time\ at\ risk}$$

$$\text{IR} = \frac{1+1}{(20\ workers)X\ (1\ day\ each)}$$



Incidence Rate (IR) – Motivating Example

- **Let's look at incidence rate instead:**

Period	# Person-day at risk	# of injuries	Injury rate
First	20	2	1 / 10 person-day
Second	20	2	1 / 10 person-day

Incidence Rate (IR) – Example 2

- The purpose of this study was to characterize suicide rates among patients with cancer in the United States. Patients in the study were residents of geographic areas served by the Surveillance, Epidemiology, and End Results (SEER) program who were diagnosed with cancer from 1973 to 2002. Among 3,594,750 SEER registry patients observed for 18,604,308 person-years, 5,838 suicides were identified. (Misono, 2008)

- **What is the incidence rate of suicide in this population?**

Incidence Rate (IR) – Example 2

- The purpose of this study was to characterize suicide rates among patients with cancer in the United States. Patients in the study were residents of geographic areas served by the Surveillance, Epidemiology, and End Results (SEER) program who were diagnosed with cancer from 1973 to 2002. Among 3,594,750 SEER registry patients observed for 18,604,308 person-years, 5,838 suicides were identified. (Misono, 2008)

- **What is the incidence rate of suicide in this population?**

$$\text{IR} = \frac{5{,}838}{18{,}604{,}308} = 0.00031$$

The incidence rate of suicide is 31 per 100,000 person-years

Mortality and Case Fatality

- **Mortality**: proportion of the population dying from the disease
  - Measuring the burden of fatal disease on the population as a whole

$$\text{Cumulative mortality} = \frac{\#\ of\ deaths\ from\ disease\ in\ a\ specific\ time\ period}{\#\ of\ disease-free\ individuals\ at\ the\ start\ of\ the\ period}$$

$$\text{Mortality rate} = \frac{\#\ of\ deaths\ from\ disease\ in\ a\ specific\ time\ period}{\#\ of\ person-time\ when\ people\ were\ at\ risk\ of\ getting\ disease}$$

- **Case fatality**: proportion of cases dying from the disease
  - Measuring the killing power of a disease
  - Typically used in acute infectious diseases such as pneumonia, ebola, COVID-19

$$\text{Case fatality} = \frac{\#\ of\ deaths\ from\ disease\ in\ a\ specific\ time\ period}{\#\ of\ individuals\ with\ the\ disease}$$

Practice Example

- City A has 200,000 residents. 400 of them had disease X. There were 1,000 deaths in one year. Of those 1000 deaths, 25 died from disease X.

- **What is the annual prevalence of disease X?**
- **What is the annual mortality?**
- **What is the annual mortality from disease X?**
- **What is the case fatality of disease X?**
- **What is the proportion of deaths from disease X?**

Practice Example

- City A has 200,000 residents. 400 of them had disease X. There were 1,000 deaths in one year. Of those 1000 deaths, 25 died from disease X.

- **What is the annual prevalence of disease X?** 400/200,000
- **What is the annual mortality?** 1,000/200,000
- **What is the annual mortality from disease X?** 25/200,000
- **What is the case fatality of disease X?** 25/400
- **What is the proportion of deaths from disease X?** 25/1,000

Odds

- Odds are a function of proportion, and are often reported in public health findings, so it is important to be familiar with them

$$\textbf{Odds} = \frac{\boldsymbol{proportion}}{\mathbf{1} - \boldsymbol{proportion}}$$

- The odds can be a number between 0 and infinity

Odds

Prevalence odds

- Providing a way to relate prevalence with incidence rate and disease duration.

$$\textbf{Prevalence odds} = \frac{prevalence}{1 - prevalence} = IR * duration$$

Cumulative incidence odds

- Commonly used in case-control studies (more in the next module)

$$\textbf{CI odds} = \frac{CI}{1 - CI}$$

Odds - Example

- In the previous example of major depressive episode prevalence in South Africa, we found that the prevalence of lifetime major depressive episode was 0.10.

- **What are the prevalence odds of lifetime major depressive episode**?

Odds - Example

- In the previous example of major depressive episode prevalence in South Africa, we found that the prevalence of lifetime major depressive episode was 0.10.

- **What are the prevalence odds of lifetime major depressive episode**?

$$\text{Prevalence odds} = \frac{0.1}{0.9} = 0.11$$

Measures of Disease Occurrence

- Prevalence
- Incidence: cumulative incidence (CI) and incidence rate (IR)
- Mortality and case fatality
- Odds

Module 1: Introduction and Measures in Epidemiologic Studies

1. What is Epidemiology?
2. Epidemiological Concepts
3. **Measures of Disease Occurrence**
4. Measures of Association
5. Measures of Impact