Introduction to Biostatistics

STAT 337

Fall 2021 (1219)*

Cameron Roopnarine† Cecilia Cotton‡

4th November 2021

*Online Course

†LaTeXer

‡Instructor

Contents

0 Module 0: Introduction to Biostatistics

Course Goal Understand common epidemiological study designs and basic biostatistical analysis methods that can be used to answer questions in health research.

Every day we make decisions that affect our health and wellness[1]

- **Exercise lowers risk of depression at all ages, researchers find.**
  - 150 minutes of activity each week is beneficial, but doing less still has positive effects.
    Amina Zafar · CBC News · Posted: 24th April 2018 | Last Updated: 24th April 2018.
- **Fewer hospital stays for asthma reported for Canadian children and teens.**
  - Research says more than half with asthma don't have it under control.
    CBC News · Posted: 26th April 2018 | Last Updated: 26th April 2018.
- **Prescription to slow worsening myopia in Canadian kids? Head outdoors.**
  - Nearly 130 % of children 11 to 13 are near-sighted, study finds.
    CBC News · Posted: 21st April 2018 | Last Updated: 23rd April 2018.
- **Opioid-related deaths nearly tripled in Ontario from 2000-2015.**
  - It's time 'to get past the stigma of drug use being among addicts,' scientist says.
    The Canadian Press · Posted: 27th April 2018 | Last Updated: 27th April 2018.
- **EU member states urged to develop co-ordinated vaccine plans for measles, flu, and other diseases.**
  - Several EU nations are facing unprecedented outbreaks of measles — a highly contagious disease that can kill.
    Thomson Reuters · Posted: 26th April 2018 | Last Updated: 26th April 2018.
- **Lung cancer patients live longer with immune therapy, study suggests.**
  - Immune therapy treatments worked for only about half of patients, but that's far better than chemo has done.
    The Associated Press · Posted: 16th April 2018 | Last Updated: 16th April 2018.

Formulating a Research Question[2,3]

Lung cancer patients live longer with immune therapy, study suggests. Immune therapy treatments worked for only about half of patients, but that's far better than chemo has done.
The Associated Press · Posted: 16th April 2018 | Last Updated: 16th April 2018.

- **Population**: Patients diagnosed with lung cancer (advanced non-small-cell lung cancer with no previous treatment).
- **Exposure**: Treatment with Immune therapy (with chemo vs chemo alone).
- **Outcome**: Overall survival.
- **Timeframe**: One year following diagnosis (study was conducted February 2016 to March 2017).

[1] Headlines taken from CBC on 30th April 2018.
[2] News Article: https://www.cbc.ca/news/health/keytruda-1.4621895
[3] Original paper: https://www.nejm.org/doi/full/10.1056/NEJMoa1801005

Example: Electronic cigarette use and smoking

Research Question: Is e-cigarette usage in youth associated with the initiation of cigarette smoking?

- Consider the following paper recently published by UW researchers.
- The full text of the paper is available through the course e-reserves.

Electronic cigarette use and smoking initiation among youth: a longitudinal cohort study

David Hammond PhD, Jessica L. Reid MSc, Adam G. Cole MSc, Scott T. Leatherdale PhD

■ Cite as: *CMAJ* 2017 October 30;189:E1328-36. doi: 10.1503/cmaj.161002

ABSTRACT

BACKGROUND: The influence of e-cigarette use on smoking initiation is a highly controversial issue, with limited longitudinal data available for examining temporal associations. We examined e-cigarette use and its association with cigarette-smoking initiation at 1-year follow-up within a large cohort of Canadian secondary school students.

METHODS: We analyzed data from students in grades 9–12 who participated in 2 waves of COMPASS, a cohort study of purposefully sampled secondary schools in Ontario and Alberta, Canada, at baseline (2013/14) and 1-year follow-up (2014/15). We assessed cigarette smoking and e-cigarette use at baseline and follow-up using self-completed surveys. We used generalized linear mixed-effects models to examine correlates of past 30-day e-cigarette use at baseline and smoking initiation between waves within the longitudinal sample.

RESULTS: Past 30-day e-cigarette use increased from 2013/14 to 2014/15 (7.2% v. 9.7%, $p < 0.001$), whereas past 30-day cigarette smoking decreased slightly (11.4% v. 10.8%, $p = 0.02$). Among the 44 163 students evaluated at baseline, past 30-day e-cigarette use was strongly associated with smoking status and smoking susceptibility. In the longitudinal sample ($n = 19\,130$), past 30-day use of e-cigarettes at baseline was associated with initiation of smoking a whole cigarette (adjusted odds ratio [OR] 2.12, 95% confidence interval [CI] 1.68–2.66) and with initiation of daily smoking (adjusted OR 1.79, 95% CI 1.41–2.28) at follow-up.

INTERPRETATION: E-cigarette use was strongly associated with cigarette smoking behaviour, including smoking initiation at follow-up. The causal nature of this association remains unclear, because common factors underlying the use of both e-cigarettes and conventional cigarettes may also account for the temporal order of initiation.

- **Population**: Canadian secondary school students (age 15–19, grade 9–12 in Alberta and Ontario).
- **Exposure**: E-cigarette usage at baseline (2013/14 school year).
- **Outcome**: Cigarette smoking initiation 1-year later (by 2014/15 school year).
- **Timeframe**: One year of follow-up.

Table 3: Between-wave smoking initiation* in the COMPASS follow-up (2014/15) longitudinal sample, Ontario and Alberta, Canada (n = 17 318)			
Characteristic (at baseline)	No. (%) within category who initiated smoking by follow-up	OR (95% CI)	
		Unadjusted†	Adjusted‡
Age, yr			
≤ 14	436/5936 (7.3)	1.0 (ref)	1.0 (ref)
15	511/5845 (8.7)	1.23 (1.07–1.41)	1.08 (0.94–1.25)
16	397/4464 (8.9)	1.27 (1.10–1.47)	1.04 (0.89–1.21)
≥ 17§	105/1073 (9.8)	1.43 (1.13–1.79)	1.02 (0.80–1.32)
Sex			
Female	683/9289 (7.4)	1.0 (ref)	1.0 (ref)
Male	766/8029 (9.5)	1.37 (1.22–1.52)	1.44 (1.28–1.62)
Past 30-day e-cigarette use			
No	1313/16 831 (7.8)	1.0 (ref)	1.0 (ref)
Yes	136/487 (27.9)	4.81 (3.90–5.94)	2.12 (1.68–2.66)

Note: CI = confidence interval, OR = odds ratio.
*Among those who had never smoked a whole cigarette at baseline, but had smoked a whole cigarette at follow-up; n = 1992 had already initiated smoking at baseline and were excluded from this measure.
†From separate generalized linear mixed models for initiating smoking between waves, including only the listed covariable, with school as a random effect (n = 17 318 for all models except race/ethnicity; n = 17 247 for race/ethnicity model).
‡From a generalized linear mixed model for initiating smoking between waves, including the covariables in the table, with school (n = 86) as a random effect (n = 17 247).
§Categories for age 17 and age ≥ 18 were combined because of low numbers for the latter category (n = 8).
¶Information on race/ethnicity was missing for 71 participants; these were excluded from the multivariable model.

Example Data Analysis (Module 2)

Smoking Initiation Status

		$D+$	$D-$	
E-cig Usage	$E+$	136	351	$487 \rightarrow 27.9\,\%$
	$E-$	1313	15518	$1638 \rightarrow 7.8\,\%$

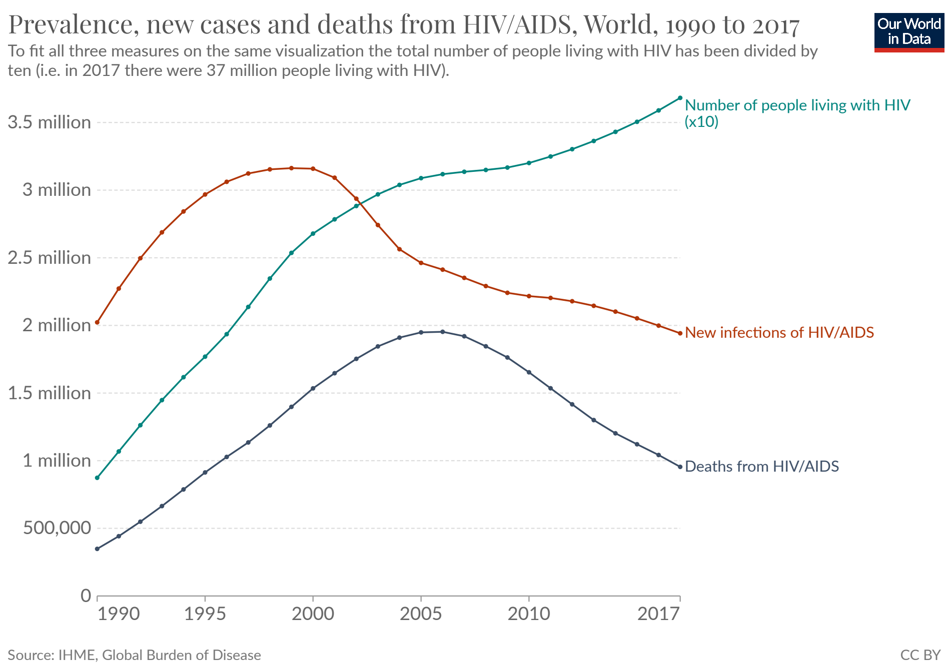
$$\text{Relative Risk} = \frac{136/487}{1313/16831} = 3.58$$

Youth who used e-cigarettes had 3.58 times the rate of smoking initiation one year later versus those who did not use e-cigarettes.

1 Module 1: Measures of Disease Frequency

1.1 Incidence and Prevalence Rates

How do we measure and evaluate patterns of disease within a population?



Primer on HIV/AIDS

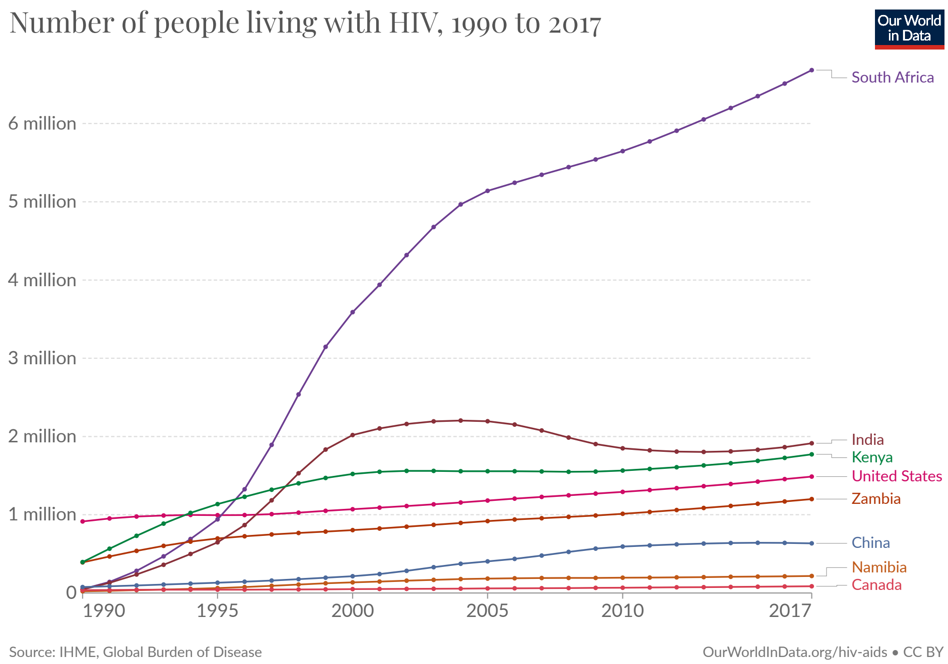
- HIV (human immunodeficiency virus) is a virus that attack's the body's immune system.
- HIV is spread through sexual contact, sharing needles, and mother-to-child transmission during pregnancy, childbirth, or breastfeeding.
- Infection with HIV can lead to AIDS (acquired immunodeficiency syndrome).
- Individuals with AIDS are at increased risk of infection and infection-related cancers.
- Currently, no cure exists, but antiretroviral therapy can slow the progression of the disease.

1.1 Incidence and Prevalence Rates

> Goal: How do we measure and evaluate patterns of disease within a population?

- Prevalence: The proportion of the population currently affected by a disease.
- Incidence: The rate at which new cases of a disease develop in a population.

Number of people living with HIV/AIDS 1990–2017



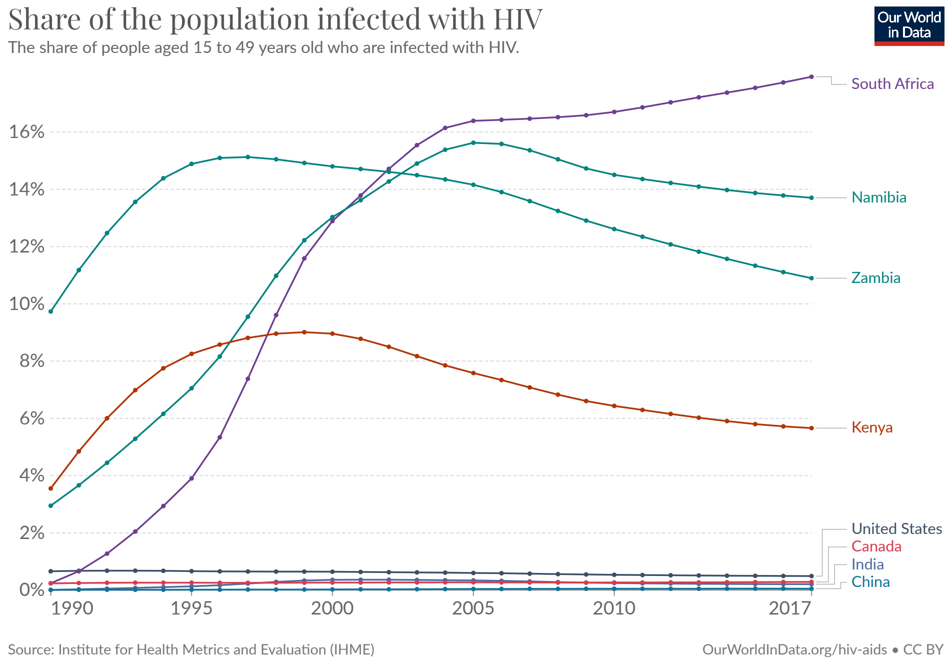
Prevalence

Prevalence: The proportion of the population currently affected by a disease.

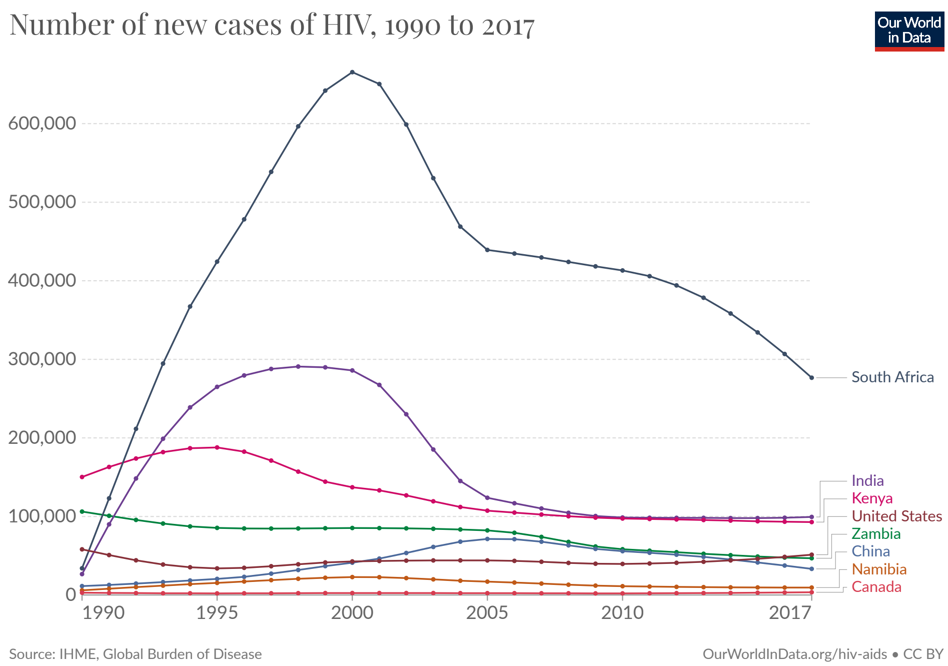
$$\text{Point Prevalence per 1000} = \frac{\text{Number of cases (new and pre-existing) in the population at a fixed point in time}}{\text{Number of individuals in the population at a fixed point in time}} \times 1000.$$

$$\text{Period Prevalence per 1000} = \frac{\text{Number of cases (new and pre-existing) in the population over a given time period}}{\text{Number of individuals in the population over a given time period}} \times 1000.$$

Prevalence of HIV/AIDS 1990–2017



Annual new cases of HIV infection, 1990–2017



Cumulative Incidence

Incidence: The rate at which **new cases** of a disease develop in a **population** over a specific **time** period.

$$\textbf{Cumulative Incidence per 1000} = \frac{\textbf{Number of new cases in the population over the time period of interest}}{\textbf{Number of individuals at risk in the population at the start of the time period of interest}} \times 1000.$$

- Assumes all subjects remain in the population and at risk for the entire time period.
- Easily violated: Births, deaths, immigration, emigration, case diagnosis.
- Consider two ways to refine the denominator calculation.
    1. Use a mid-interval population estimate.
    2. Calculate the total person-time at risk in the population.

Incidence Density or Incidence Rate

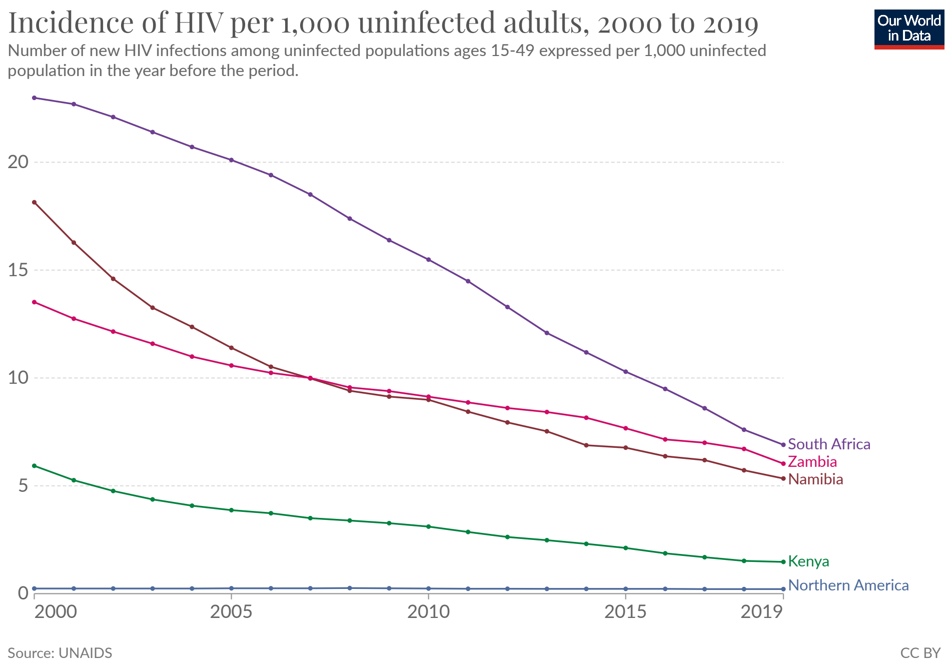
$$\textbf{Incidence Density per 1000} = \frac{\textbf{Number of new cases in the population over the time period of interest}}{\textbf{Mid-interval estimate of the population}} \times 1000.$$

- 3,218 new cases of HIV in Canada, 2016.
- 36,264,604 July 1, 2016 Canadian population estimate.

$$\textbf{Incidence Density per 100,000} = \frac{3,218}{36,264,604} \times 100,000 = 8.873666.$$

- *The incidence of HIV infection in Canada in the year 2016 was 8.87 cases per 100,000 persons*.

Incidence of HIV per 1,000 uninfected adults, 2000–2017



Person-time at risk

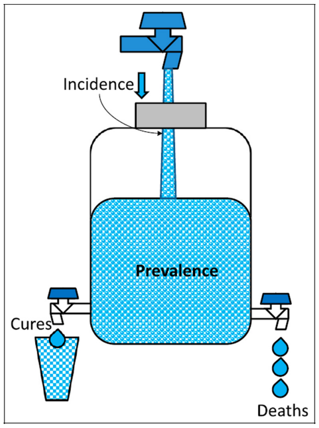
- To account for varying time periods of risk we consider an alternative denominator for our incidence calculation.
- Person-time at risk is the duration of time an individual is at risk for developing a disease.
- Assuming they are initially disease free, it is the length of time from baseline until the first of:
  1. They develop the disease of interest and become a case.
  2. They cease to be at risk of becoming a case due to either death from unrelated causes or they leave the population.
  3. The end of the time period of interest is reached.
- Total person-time at risk is the sum of the individual contributions over the population.

Incidence Density or Incidence Rate

$$\begin{array}{c}\textbf{Incidence Density}\\ \textbf{per 1000}\end{array} = \frac{\begin{array}{c}\textbf{Number of new cases in the}\\ \textbf{population over the time period of interest}\end{array}}{\begin{array}{c}\textbf{Mid-Total person-time at risk in the}\\ \textbf{population over the time period of interest}\end{array}} \times 1000.$$

- Incidence density estimate is more precise than cumulative incidence, but may be harder to get information needed, so this measure is often used for small populations.
- Expressed as per 10^x person-years (-month, -day).

Relationship Between Incidence and Prevalence



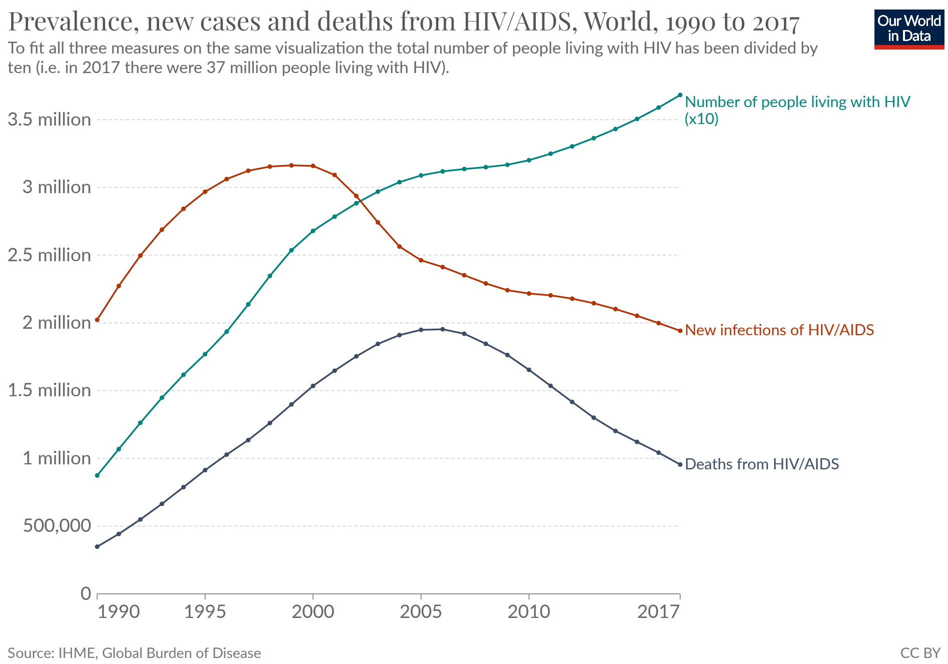
Relationship Between Incidence and Prevalence

- **Incidence**: the rate new cases are diagnosed in a population.
- **Prevalence**: the proportion of the population currently affected by the disease.

$$\textbf{Prevalence} \approx \textbf{Incidence} \times \textbf{Disease Duration}$$

- Relationship is approximate but generally holds well if prevalence is low ($< 10\,\%$) and duration is fairly constant (or an average can be taken).
- Note: units must be consistent in order to perform the multiplication operation.

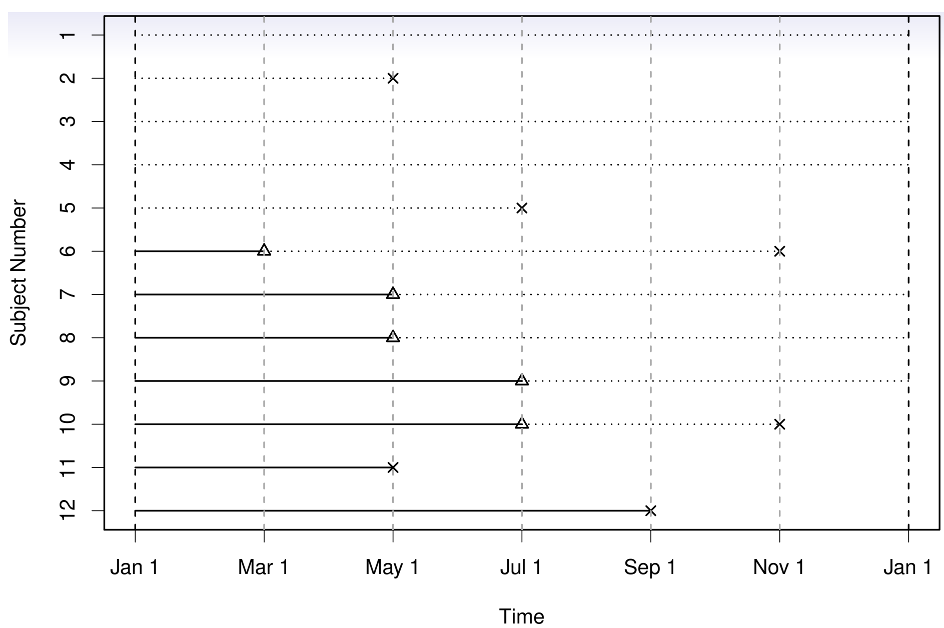
Prevalence, new cases, and mortality for HIV/AIDS



Exercise: Incidence and Prevalence Calculations

Total population size of 100. Histories of 12 subjects with disease are below. Subjects 13–100 do not have the disease during the year of study. (△ Diagnosis; × Death)

Subject	Diagnosis △	Death ×
1	< January 1	
2	< January 1	April 30
3	< January 1	
4	< January 1	
5	< January 1	June 30
6	March 1	October 31
7	May 1	
8	May 1	
9	July 1	
10	July 1	October 31
11	NA	May 1
12	NA	September 1
13–100	NA	



- Point Prevalence on July 1.
- Period Prevalence (Jan 1 to Dec 31)
- Cumulative Incidence (Jan 1 to Dec 31).
- Incidence Density (Jan 1 to Dec 31).

$$\begin{aligned}
\begin{array}{c}\text{Point Prevalence}\\ \text{on July 1}\end{array} &= \frac{\text{\# cases in the pop on July 1}}{\text{\# indv in the pop on July 1}} \times 1000 = \frac{8}{97} \times 1000 = 82.47 \text{ per 1000 persons.}\\
\begin{array}{c}\text{Period Prevalence}\\ \text{(Jan 1-Dec 31)}\end{array} &= \frac{\text{\# cases during Jan 1-Dec 31}}{\text{\# indv in the pop Jan 1-Dec 31}} \times 1000 = \frac{10}{100} \times 1000 = 100 \text{ per 1000 persons.}\\
\begin{array}{c}\text{Cumulative Incidence}\\ \text{(Jan 1-Dec 31)}\end{array} &= \frac{\text{\# new cases during Jan 1-Dec 31}}{\text{\# indv at risk on Jan 1}} \times 1000 = \frac{5}{95} \times 1000 = 52.63 \text{ per 1000 persons.}\\
\begin{array}{c}\text{Incidence Density}\\ \text{(Jan 1-Dec 31)}\end{array} &= \frac{\text{\# new cases during Jan 1-Dec 31}}{\text{July 1 population size}} \times 1000 = \frac{5}{97} \times 1000 = 51.55 \text{ per 1000 persons.}\\
\begin{array}{c}\text{Incidence Density}\\ \text{(Jan 1-Dec 31)}\end{array} &= \frac{\text{\# new cases during Jan 1-Dec 31}}{\text{Total person-years at risk Jan 1-Dec 31}} \times 1000 = \frac{5}{90.83} \times 1000 = 55.05 \text{ per 1000 p-y.}
\end{aligned}$$

1.2 Standardization of Rates: Indirect Methods