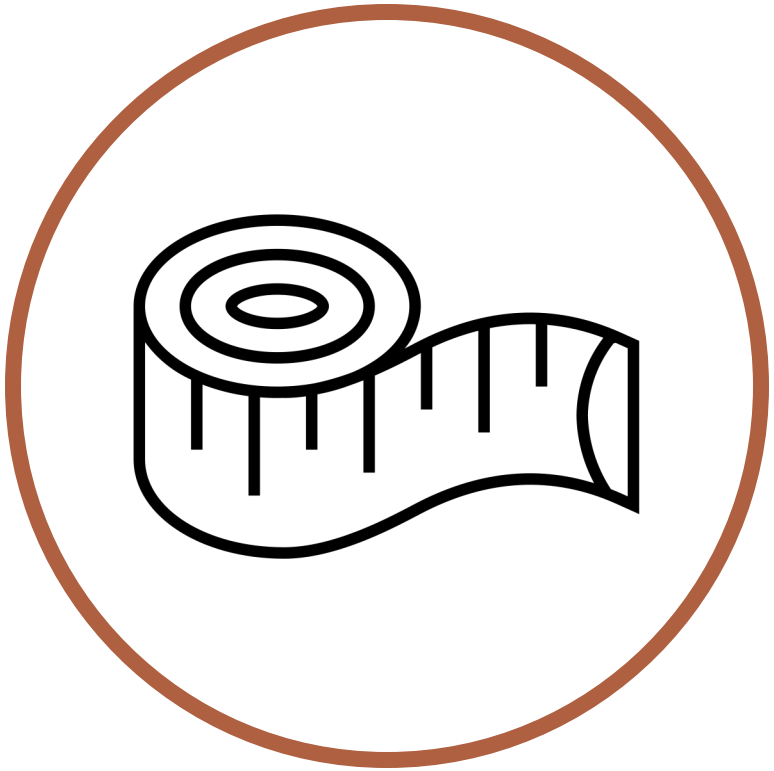
Measures of Association

Epidemiology III

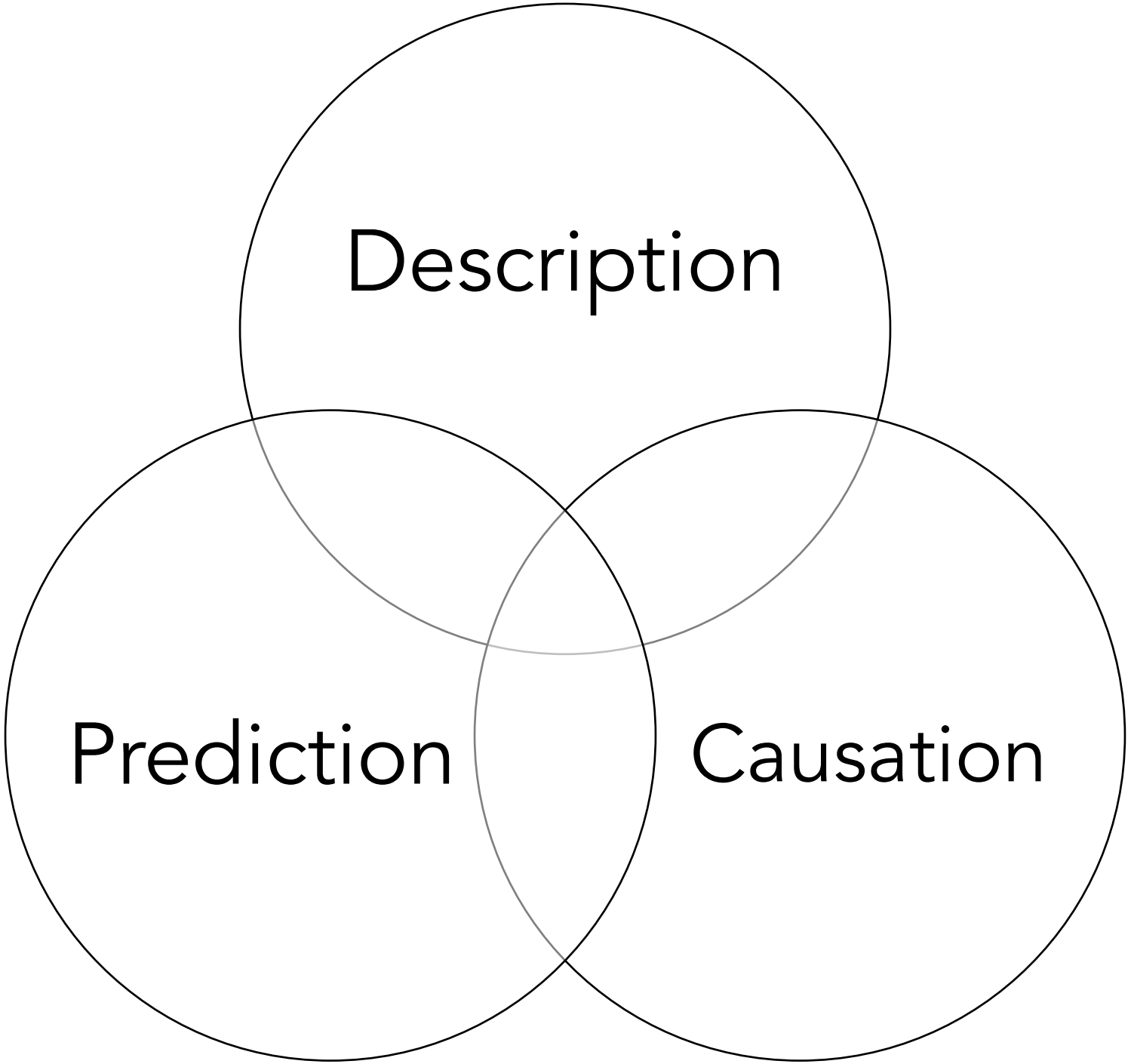


Measurement

"the study of the occurrence and distribution of health-related states or events in specified populations, including the study of the determinants influencing such states, and the application of this knowledge to control the health problems."

- Observe and assign values to relevant characteristics.
- Look for patterns among those values.

Source: Porta M, ed. *A Dictionary of Epidemiology*. Oxford University Press; 2008.


Description
Prediction
Causation

Description

- Not necessarily looking for associations.
- Distributions (i.e., middle, spread, shape, count, proportion) of single variables.
- Examples:
  - How many ventilators are available in Texas?
  - What is the average age of people living in Florida?
  - How much time elapses, on average, between exposure to a pathogen and occurrence of disease symptoms.

Description

- But, sometimes looking for associations.
- Comparing distributions of single variables within levels of another variable.
- Examples:
  - Are there more ventilators available in Texas or New York?
  - Are people older, on average, in Florida or Pennsylvania?
  - Is symptom onset quicker, on average, for Cholera or E. coli?

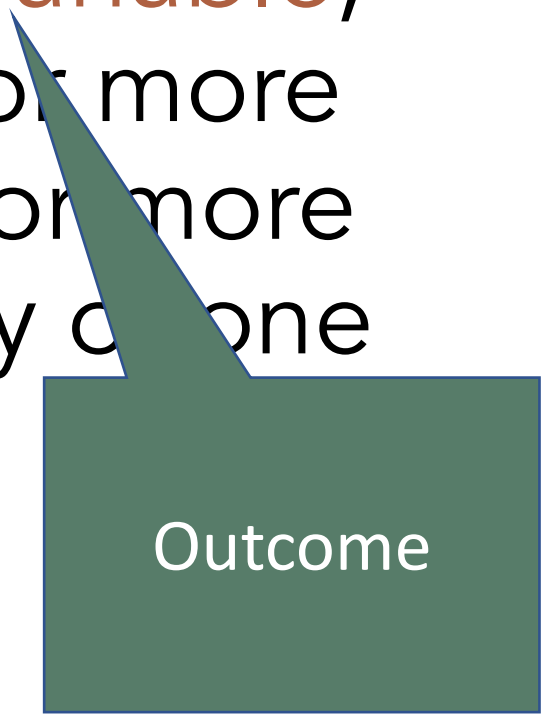
Prediction

- Implies associations.

"[An association is a] statistical dependence between two or more events, characteristics, or other variables. An association is present if the probability of occurrence of an event or characteristic, or the quantity of a variable, varies with the occurrence of one or more other events, the presence of one or more other characteristics, or the quantity of one or more other variables."

Source: Porta, Miquel. A Dictionary of Epidemiology (pp. 8-9). Oxford University Press. Kindle Edition.

"[An association is a] statistical dependence between two or more events, characteristics, or other variables. An association is present if the probability of occurrence of an event or characteristic, or the quantity of a variable, varies with the occurrence of one or more other events, the presence of one or more other characteristics, or the quantity of one or more other variables."



Source: Porta, Miquel. A Dictionary of Epidemiology (pp. 8-9). Oxford University Press. Kindle Edition.

"[An association is a] statistical dependence between two or more events, characteristics, or other variables. An association is present if the probability of occurrence of an event or characteristic, or the quantity of a variable, varies with the occurrence of one or more other events, the presence of one or more other characteristics, or the quantity of one or more other variables."

Exposure(s)

Source: Porta, Miquel. A Dictionary of Epidemiology (pp. 8-9). Oxford University Press. Kindle Edition.

Exposure

"In epidemiology, the exposed group (or simply, the exposed) is often used to connote a group whose members have been exposed to a supposed cause of a disease or health state of interest or possess a characteristic that is a determinant of the health outcome of interest."

Source: Porta, Miquel. A Dictionary of Epidemiology (pp. 8-9). Oxford University Press. Kindle Edition.

Exposure

- A pathogen (e.g., HPV)
- An environmental condition (e.g., lead paint)
- A Behavior (e.g., regular physical activity)
- A trait (e.g., APOE ε4 allele)
- A social condition (e.g., discrimination)
- A comorbid disease or health condition (e.g., diabetes)
- Treatment or intervention (e.g., smoking cessation program)

Source: https://www.merriam-webster.com/dictionary/exposure

Association

$$\Pr[Y = 1|A = 1] \neq \Pr[Y = 1|A = 0]$$

Source: Hernán MA, Robins JM. *Causal Inference: What If*. CRC Press; 2020.

Association

$$\Pr[Y = 1|A = 1] \neq \Pr[Y = 1|A = 0]$$

- Pr[] means "the probability of"
- Y = 1 means "the outcome happens" (
  - Y = 0 means it doesn't happen
  - For example, Alzheimer's disease happens
- | means "given that" or "if"
- A = 1 means "A happens"
  - A = 0 means it doesn't happen
  - For example, APOE **ε**4 allele
- ≠ means "does not equal"

Source: Hernán MA, Robins JM. *Causal Inference: What If*. CRC Press; 2020.

Association

$$\Pr[Y = 1|A = 1] \neq \Pr[Y = 1|A = 0]$$

The probability that Y equals 1 given that A equals 1 is not equal to the probability that Y equals 1 given that A equals 0.

The probability of Alzheimer's Disease among people who carry at least one APOE **ε**4 allele is not equal to the probability of Alzheimer's Disease among people who do not carry any APOE **ε**4 alleles

Source: Hernán MA, Robins JM. *Causal Inference: What If*. CRC Press; 2020.

Probability

- How likely an event is to happen.

Probability

- How likely an event is to happen.
  - Frequency probability: The limit of the relative frequency of an event in a sequence of N random trials as N approaches infinity.
  - Subjective probability: A measure, ranging from 0 to 1, of the degree of belief in a hypothesis or statement.

Source: Porta M, ed. *A Dictionary of Epidemiology*. Oxford University Press; 2008.

Probability

- How likely an event is to happen.
  - Frequency probability: The limit of the relative frequency of an event in a sequence of N random trials as N approaches infinity.
  - Subjective probability: A measure, ranging from 0 to 1, of the degree of belief in a hypothesis or statement.

$$P(Y) = \frac{Number\ of\ times\ outcome\ Y\ occurs}{Total\ number\ of\ all\ possible\ outcomes}$$

Source: Porta M, ed. *A Dictionary of Epidemiology*. Oxford University Press; 2008.

Probability

- How likely an event is to happen.
  - Frequency probability: The limit of the relative frequency of an event in a sequence of N random trials as N approaches infinity.
  - Subjective probability: A measure, ranging from 0 to 1, of the degree of belief in a hypothesis or statement.

$$P(Y) = \frac{Number\ of\ times\ Alzheimer's\ Disease\ occurs}{Total\ number\ of\ times\ Alzheimer's\ Disease\ could\ have\ occured}$$

Source: Porta M, ed. *A Dictionary of Epidemiology*. Oxford University Press; 2008.

Probability

- How likely an event is to happen.
  - Frequency probability: The limit of the relative frequency of an event in a sequence of N random trials as N approaches infinity.
  - Subjective probability: A measure, ranging from 0 to 1, of the degree of belief in a hypothesis or statement.

Marginal or Unconditional Probability

$$P(Y) = \frac{Number\ of\ times\ outcome\ Y\ occurs}{Total\ number\ of\ all\ possible\ outcomes}$$

Source: Porta M, ed. *A Dictionary of Epidemiology*. Oxford University Press; 2008.

Conditional Probability

- How likely an event is to happen given that something else has already happened.

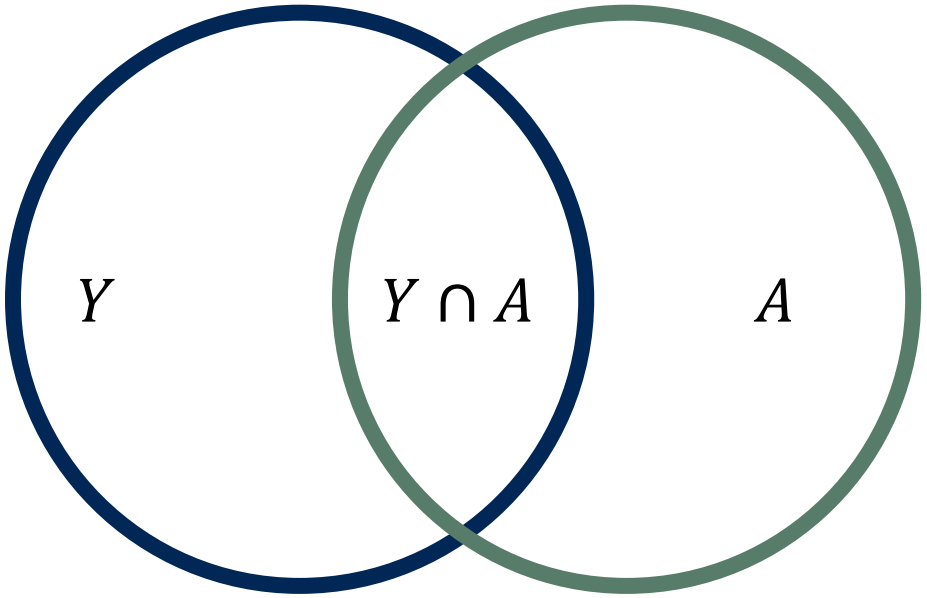
$$P(Y|A) = \frac{P(Y \cap A)}{P(A)}$$

∩ means "intersects"

Conditional Probability

- How likely an event is to happen given that something else has already happened.

$$P(Y|A) = \frac{P(Y \cap A)}{P(A)}$$



Conditional Probability

- How likely an event is to happen given that something else has already happened.

$$P(Y|A) = \frac{P(Y \cap A)}{P(A)}$$

$$P(Y|A) = \frac{Number\ of\ times\ outcome\ Y\ occurs}{Number\ of\ times\ condtion\ A\ has\ occurred}$$

Conditional Probability

- How likely an event is to happen given that something else has already happened.

$$P(Y|A) = \frac{P(Y \cap A)}{P(A)}$$

$$P(Y|A) = \frac{Number\ of\ times\ Alzheimer's\ Disease\ occurs}{People\ with\ APOE\ \varepsilon 4\ allele}$$

Conditional Probability

- How likely an event is to happen given that something else has already happened.

$$P(Y = 1|A = 1) = \frac{P(Y = 1 \cap A = 1)}{P(A = 1)}$$

$$P(Y = 1|A = 1) = \frac{Number\ of\ times\ Alzheimer's\ Disease\ occurs}{People\ with\ APOE\ \varepsilon 4\ allele}$$

Association

$$\Pr[Y = 1|A = 1] \neq \Pr[Y = 1|A = 0]$$

Source: Hernán MA, Robins JM. *Causal Inference: What If*. CRC Press; 2020.

No Association (Independence)

$$\Pr[Y = 1|A = 1] = \Pr[Y = 1|A = 0]$$

$$\Pr[Y = 1|A = 1] - \Pr[Y = 1|A = 0] = 0$$

$$\frac{\Pr[Y = 1|A = 1]}{\Pr[Y = 1|A = 0]} = 1$$

$$\frac{\Pr[Y = 1|A = 1]/\Pr[Y = 0|A = 1]}{\Pr[Y = 1|A = 0]/\Pr[Y = 0|A = 0]} = 1$$

Source: Hernán MA, Robins JM. *Causal Inference: What If*. CRC Press; 2020.

No Association (Independence)

$$\Pr[Y = 1|A = 1] = \Pr[Y = 1|A = 0]$$

No Incidence Proportion difference association

$$\Pr[Y = 1|A = 1] - \Pr[Y = 1|A = 0] = 0$$

No Incidence Proportion ratio association

$$\frac{\Pr[Y = 1|A = 1]}{\Pr[Y = 1|A = 0]} = 1$$

No Incidence odds ratio association

$$\frac{\Pr[Y = 1|A = 1]/\Pr[Y = 0|A = 1]}{\Pr[Y = 1|A = 0]/\Pr[Y = 0|A = 0]} = 1$$

Source: Hernán MA, Robins JM. *Causal Inference: What If*. CRC Press; 2020.

Null value

"The null value of a measure of association is the value that measure takes when there is no difference between the two groups being compared."

Source: Westreich D. *Epidemiology by Design*. Oxford University Press; 2019

No Association (Independence)

$$\Pr[Y = 1|A = 1] = \Pr[Y = 1|A = 0]$$

Incidence Proportion difference association

$$\Pr[Y = 1|A = 1] - \Pr[Y = 1|A = 0] = 0$$

Null = 0

Incidence Proportion ratio association

$$\frac{\Pr[Y = 1|A = 1]}{\Pr[Y = 1|A = 0]} = 1$$

Null = 1

Incidence odds ratio association

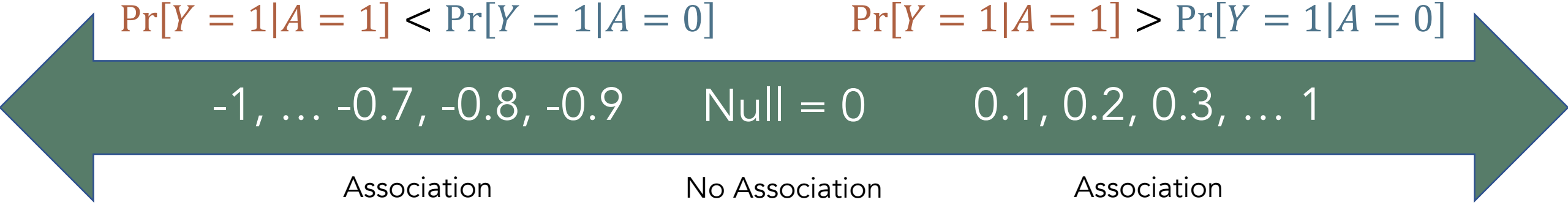
$$\frac{\Pr[Y = 1|A = 1]/\Pr[Y = 0|A = 1]}{\Pr[Y = 1|A = 0]/\Pr[Y = 0|A = 0]} = 1$$

Null = 1

Source: Hernán MA, Robins JM. *Causal Inference: What If*. CRC Press; 2020.

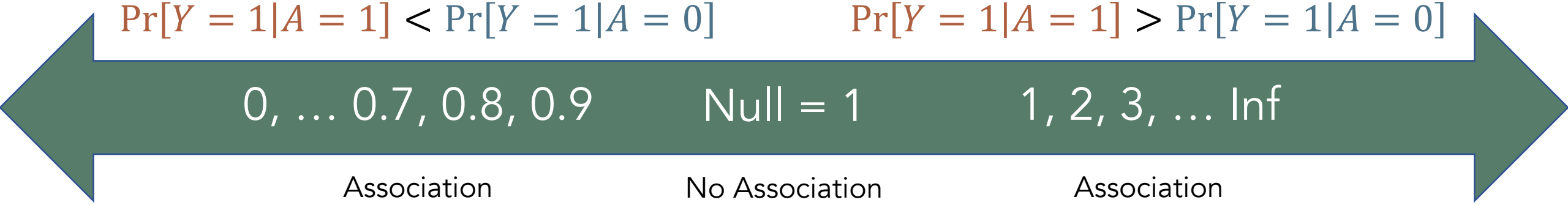
Risk Difference

$$\Pr[Y = 1|A = 1] - \Pr[Y = 1|A = 0] = \cdots$$



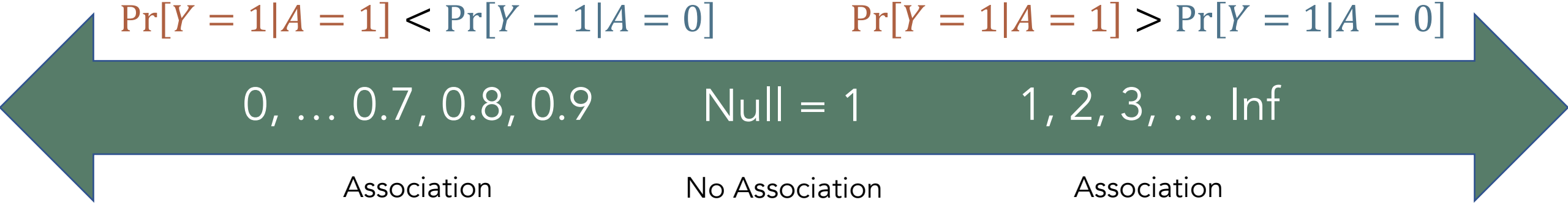
Risk Ratio

$$\frac{\Pr[Y = 1|A = 1]}{\Pr[Y = 1|A = 0]} = \dots$$



Odds Ratio

$$\frac{\Pr[Y = 1|A = 1]/\Pr[Y = 0|A = 1]}{\Pr[Y = 1|A = 0]/\Pr[Y = 0|A = 0]} = \dots$$



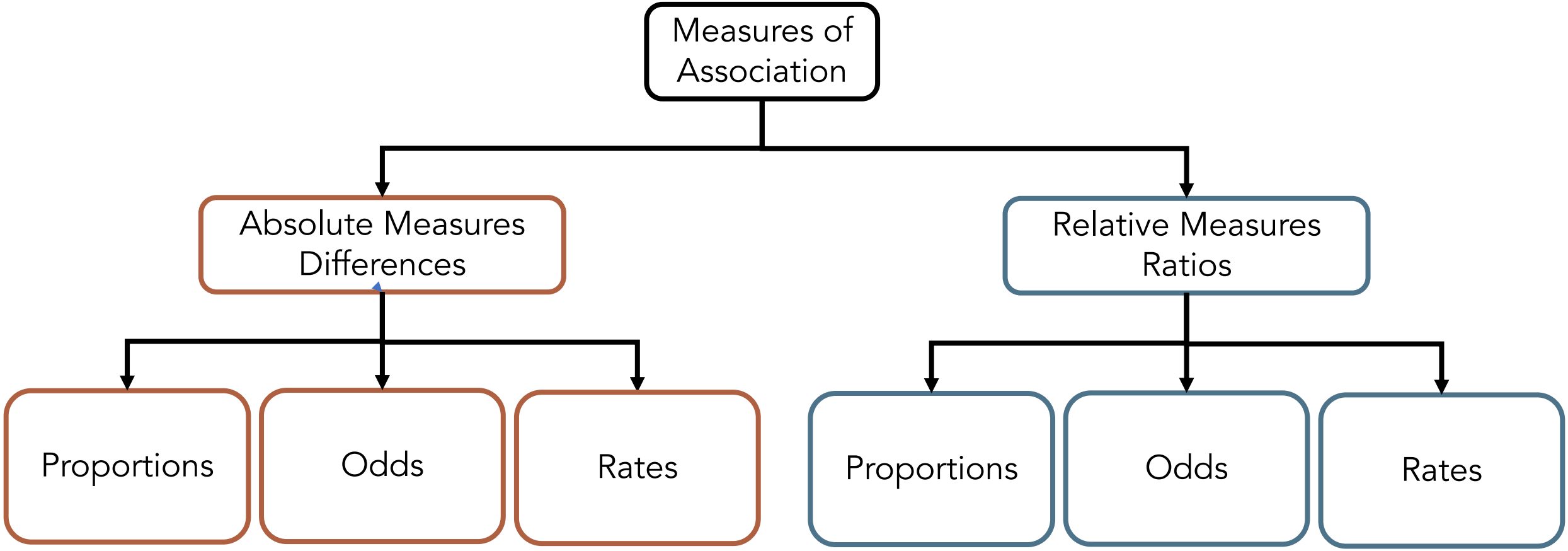
Association

- The distribution (i.e., middle, spread, shape, proportion of people in each category) of the thing we are measuring is different, on average, in two groups.
- Knowing something about X tells you something about Y.

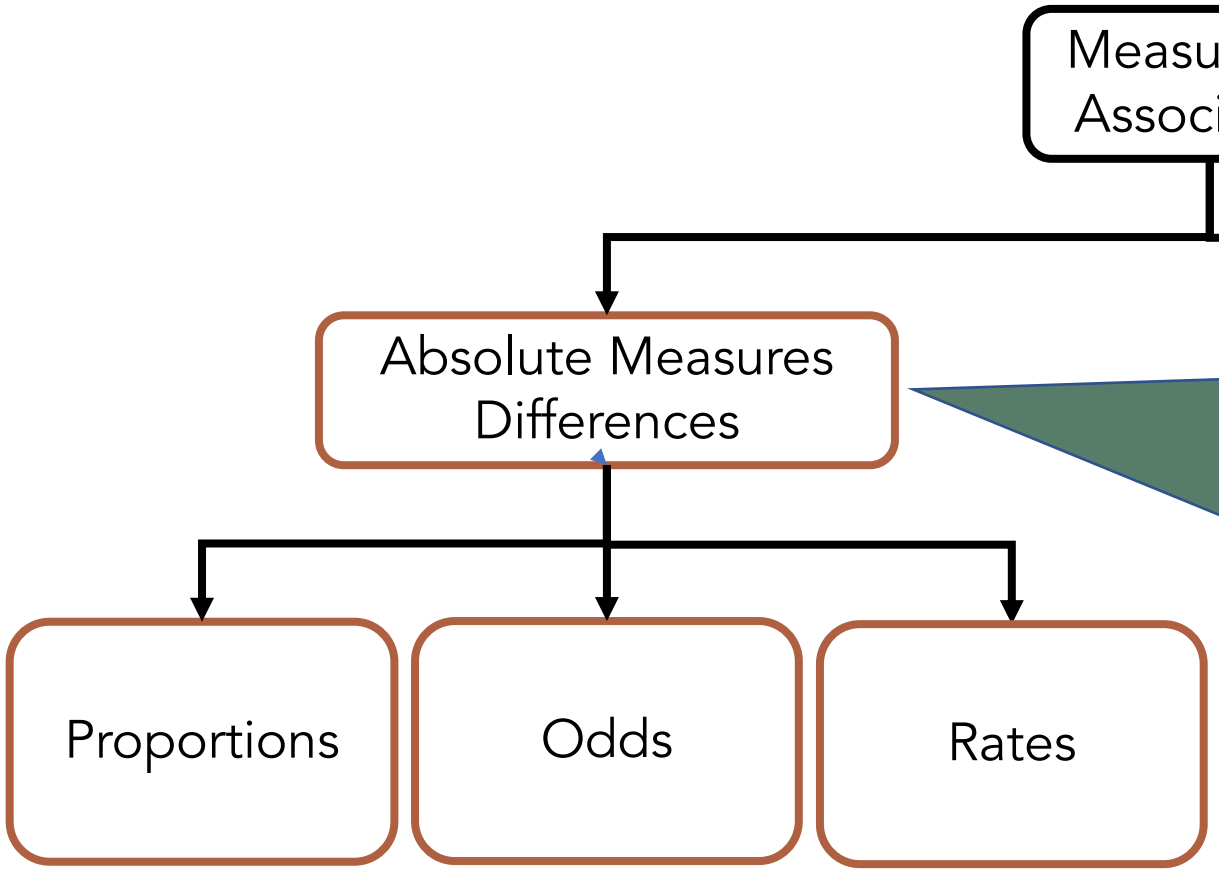
Source: Hernán MA. Causal Diagrams: Draw Your Assumptions Before Your Conclusions. edX. Accessed September 27, 2022. https://learning.edx.org/course/course-v1:HarvardX+PH559x+2T2022/home

Prediction

- Implies associations.
- Examples
  - People of a certain race/ethnicity are more likely to get a particular cancer.
  - Older adults who have trouble managing finances are more likely to get dementia.
- Can be useful on its own.
- Often, in epidemiology, the ultimate goal is causal inference.
- "the study of the occurrence and distribution of health-related states or events in specified populations, including the study of the determinants influencing such states, and the application of this knowledge to control the health problems." (Porta 2008, page 81)

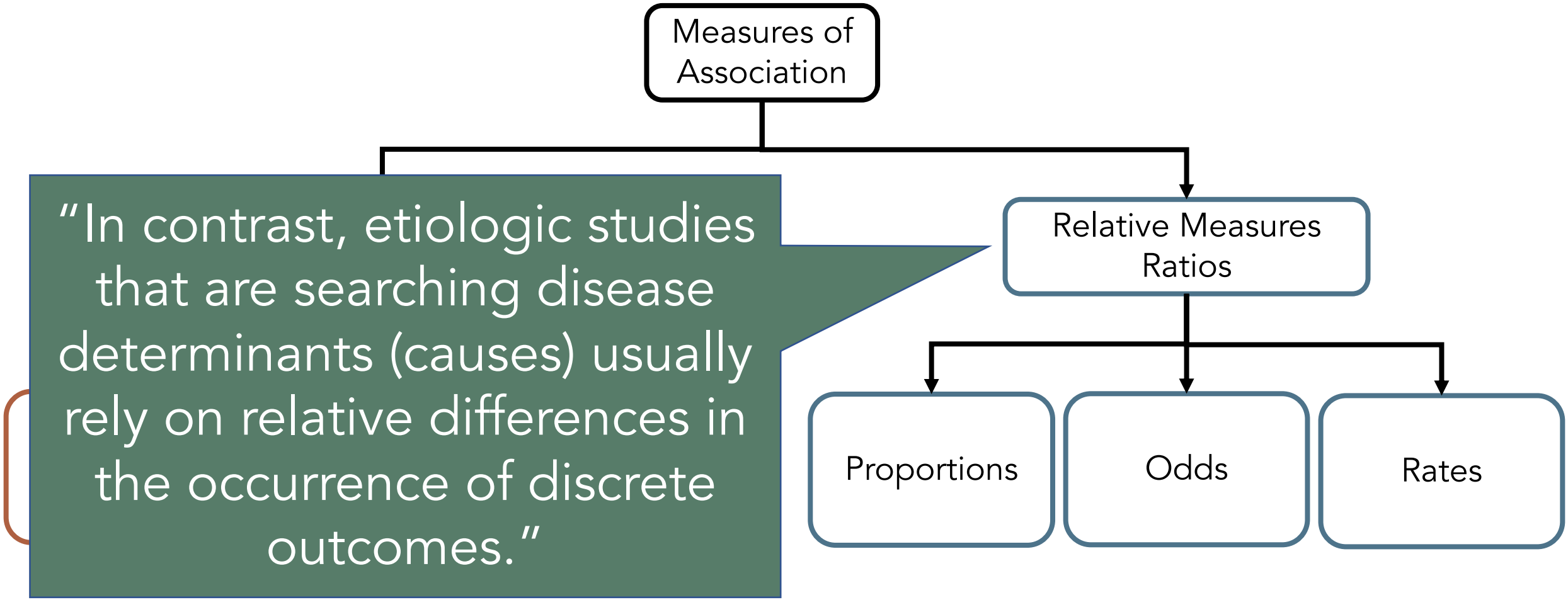


Source: Szklo M, Nieto FJ. *Epidemiology: Beyond the Basics*. Jones & Bartlett Learning; 2019.



"Measures based on absolute differences are often preferred when public health or preventive activities are evaluated, as their main goal is often an absolute reduction in the risk of an undesirable outcome."

Source: Szklo M, Nieto FJ. *Epidemiology: Beyond the Basics*. Jones & Bartlett Learning; 2019.

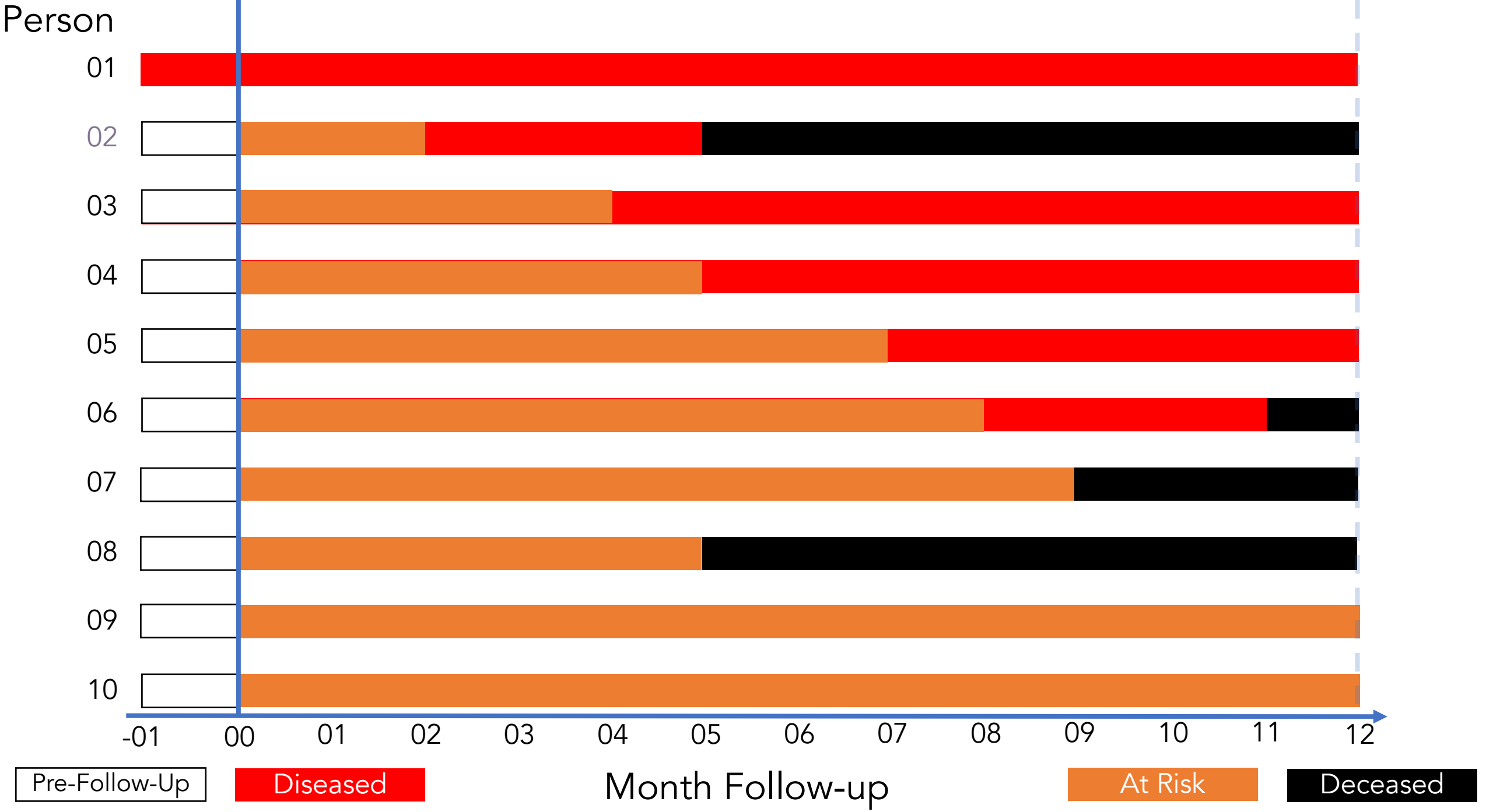


Source: Szklo M, Nieto FJ. *Epidemiology: Beyond the Basics*. Jones & Bartlett Learning; 2019.

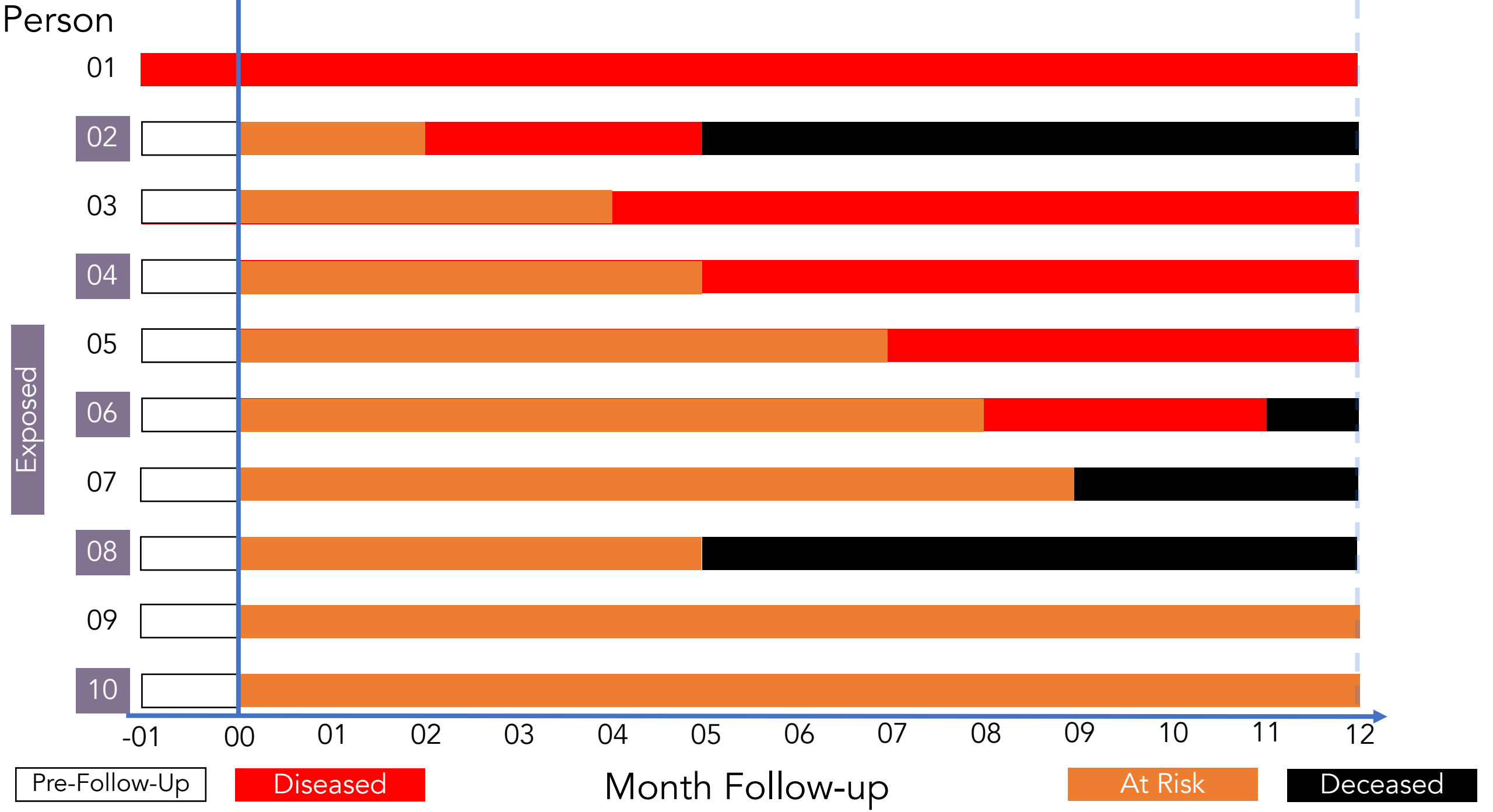
2x2 Contingency Table (Crosstab)

	Outcome +	Outcome -
Exposure +	a	b
Exposure -	c	d

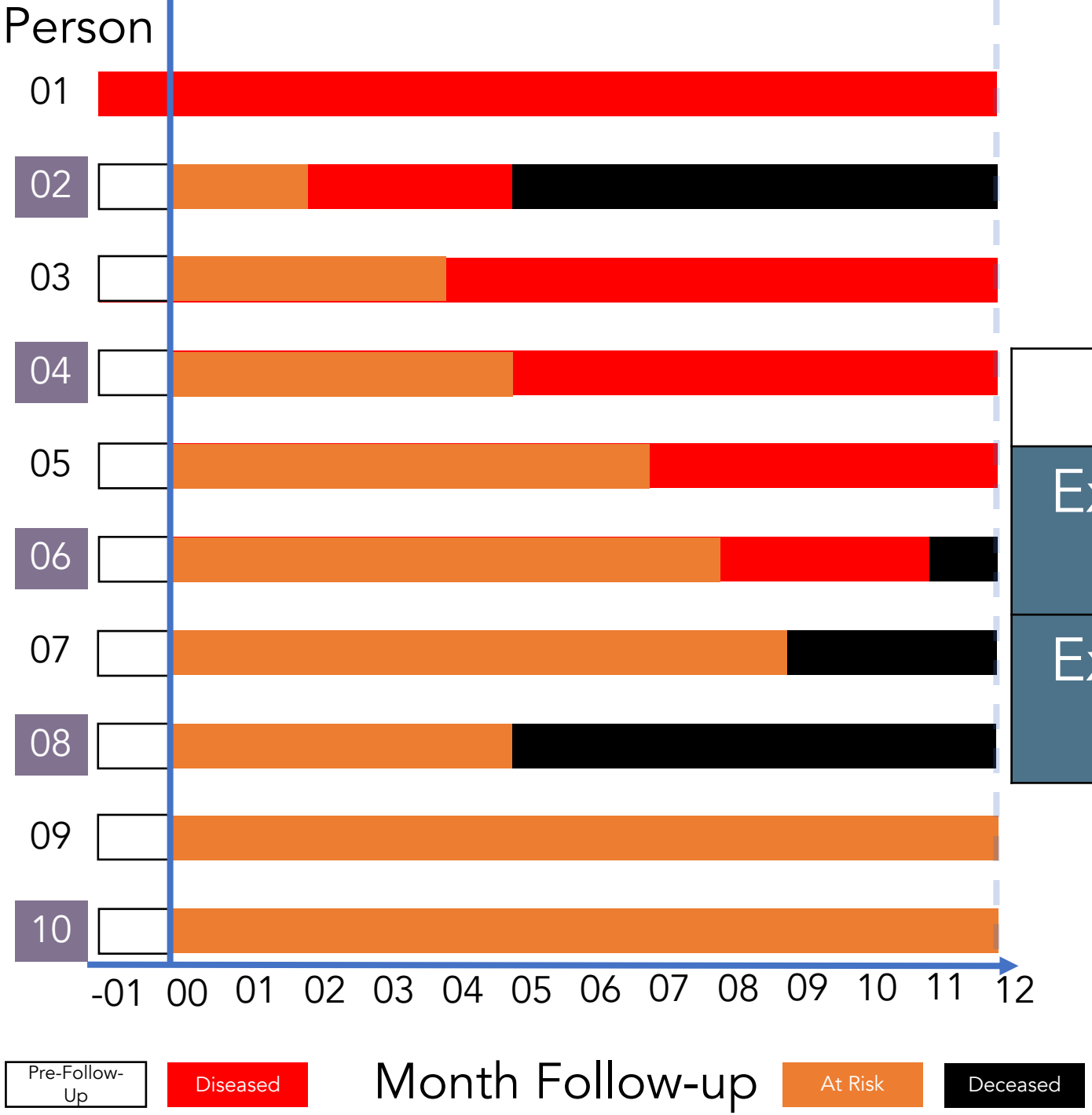
This order is important!!



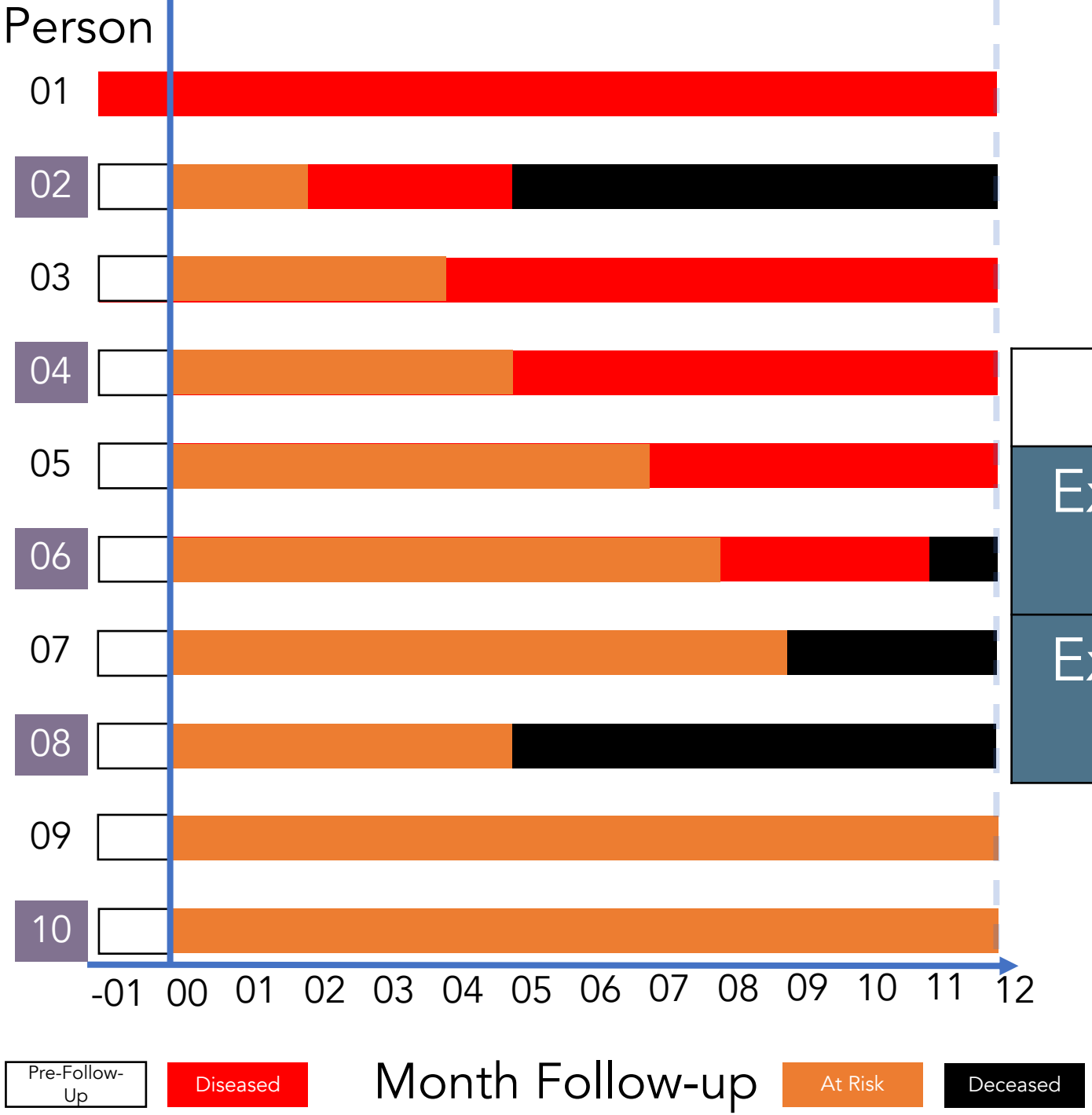
Source: Lash TL, VanderWeel TJ, Haneuse S, Rothman KJ. *Modern Epidemiology*. fourth. Wolters Kluwer; 2021.



Source: Lash TL, VanderWeel TJ, Haneuse S, Rothman KJ. *Modern Epidemiology*. fourth. Wolters Kluwer; 2021.



	Disease +	Disease -
Exposure +	3	?
Exposure -	?	?



	Disease +	Disease -
Exposure +	3	2
Exposure -	3	2

2x2 Contingency Table (Crosstab)

	Outcome +	Outcome -	Row Total
Exposure +	a	b	a+b
Exposure -	c	d	c+d
Column Total	a+c	b+d	a+b+c+d

2x2 Contingency Table (Crosstab)

	Outcome +	Outcome -	Row Total
Exposure +	a	b	a+b
Exposure -	c	d	c+d
Column Total	a+c	a+c	a+b+c+d

Table Margins
Marginal Totals

2x2 Contingency Table (Crosstab)

	Outcome +	Outcome -	Row Total
Exposure +	3	2	5
Exposure -	3	2	?
Column Total	?	?	?

2x2 Contingency Table (Crosstab)

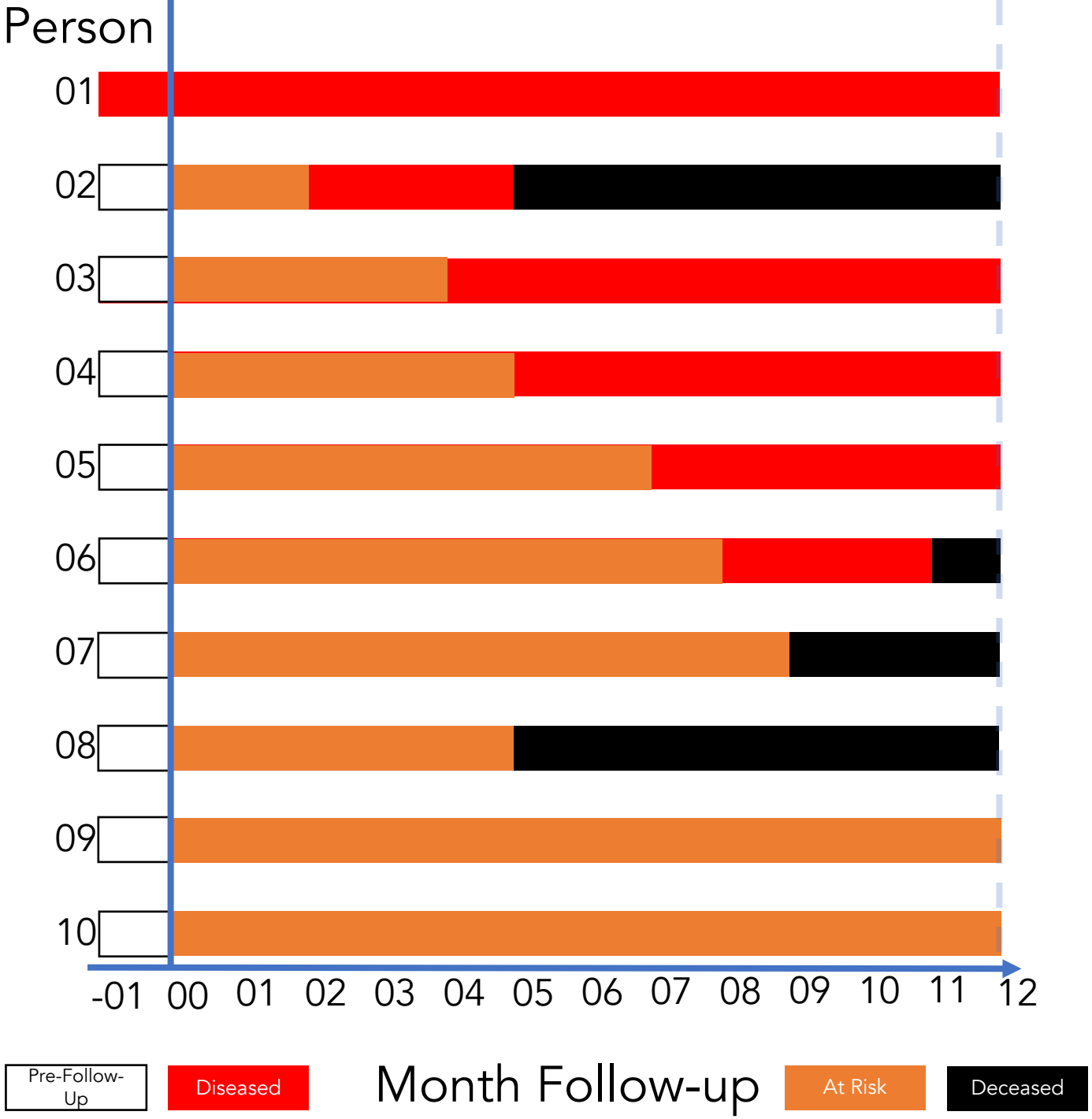
	Outcome +	Outcome -	Row Total
Exposure +	3	2	5
Exposure -	3	2	5
Column Total	6	4	10

Source: https://www.r-project.org/logo/

Incidence Proportion

- The proportion of the population who experiences a new occurrence of the condition of interest among those in the population who are at risk of experiencing a new occurrence the condition of interest during a given time frame.
- Like all proportions, falls between 0 and 1.

Source: Westreich D. *Epidemiology by Design*. Oxford University Press; 2019



Incidence Proportion

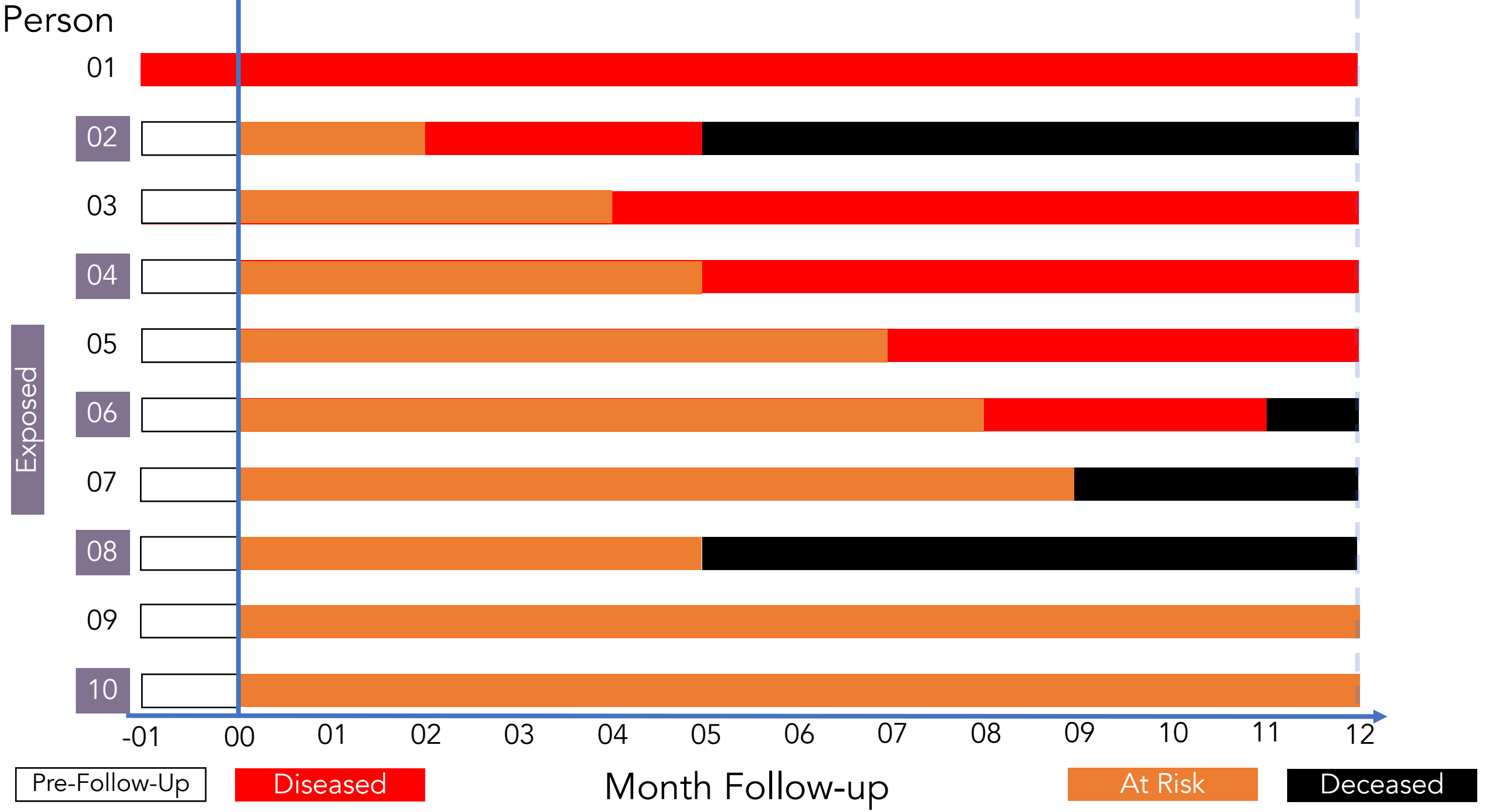
$$\frac{Count\ of\ new\ occurences}{Population\ at\ risk}$$

$=$

$\frac{5}{9}$

$=$

$0.56\ or\ 56\%$



Source: Lash TL, VanderWeel TJ, Haneuse S, Rothman KJ. *Modern Epidemiology*. fourth. Wolters Kluwer; 2021.

Incidence Proportion Ratio

- More commonly called
  - Relative risk
  - Risk ratio

Incidence Proportion Ratio

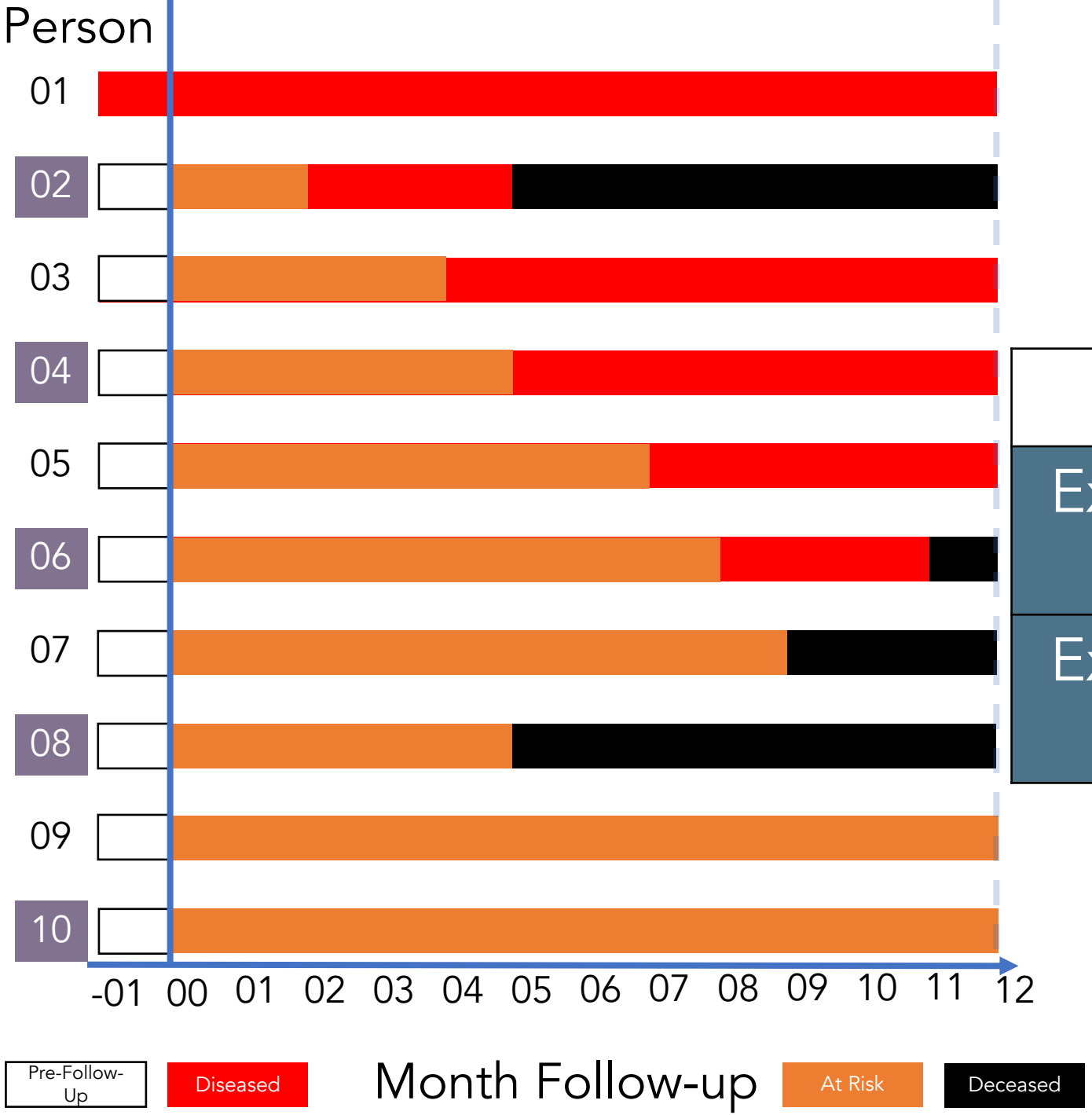
$$\frac{Incidence\ proportion\ in\ the\ exposed}{Incidence\ proportion\ in\ the\ unexposed}$$

Incidence Proportion Ratio

	Outcome +	Outcome -	Row Total
Exposure +	a	b	a+b
Exposure -	c	d	c+d
Column Total	a+c	a+c	a+b+c+d

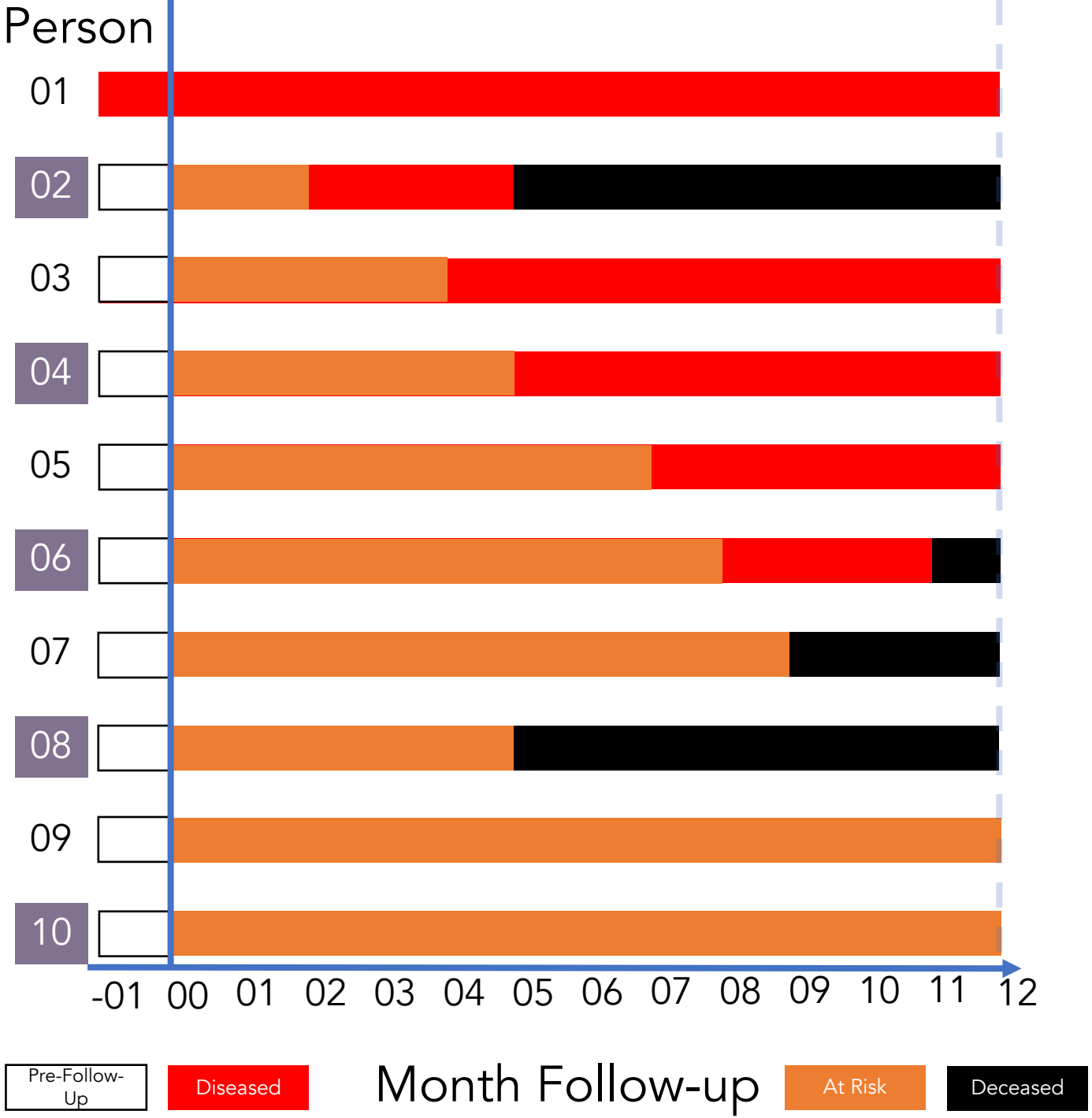
The incidence proportion among exposed = a / (a + b)
The incidence proportion among unexposed = c / (c + d)

$$\text{Incidence Proportion Ratio} = \frac{a/(a+b)}{c/(c+d)}$$



	Disease +	Disease -
Exposure +	3	2
Exposure -	2	2

Source: https://www.r-project.org/logo/



Incidence Proportion Ratio

Among the members of our population, those who were exposed had 1.2 times the incidence of disease compared to those who were not exposed over the 12 months of follow-up period.

Incidence Proportion Difference

- More commonly called
  - Risk difference

Incidence Proportion Difference

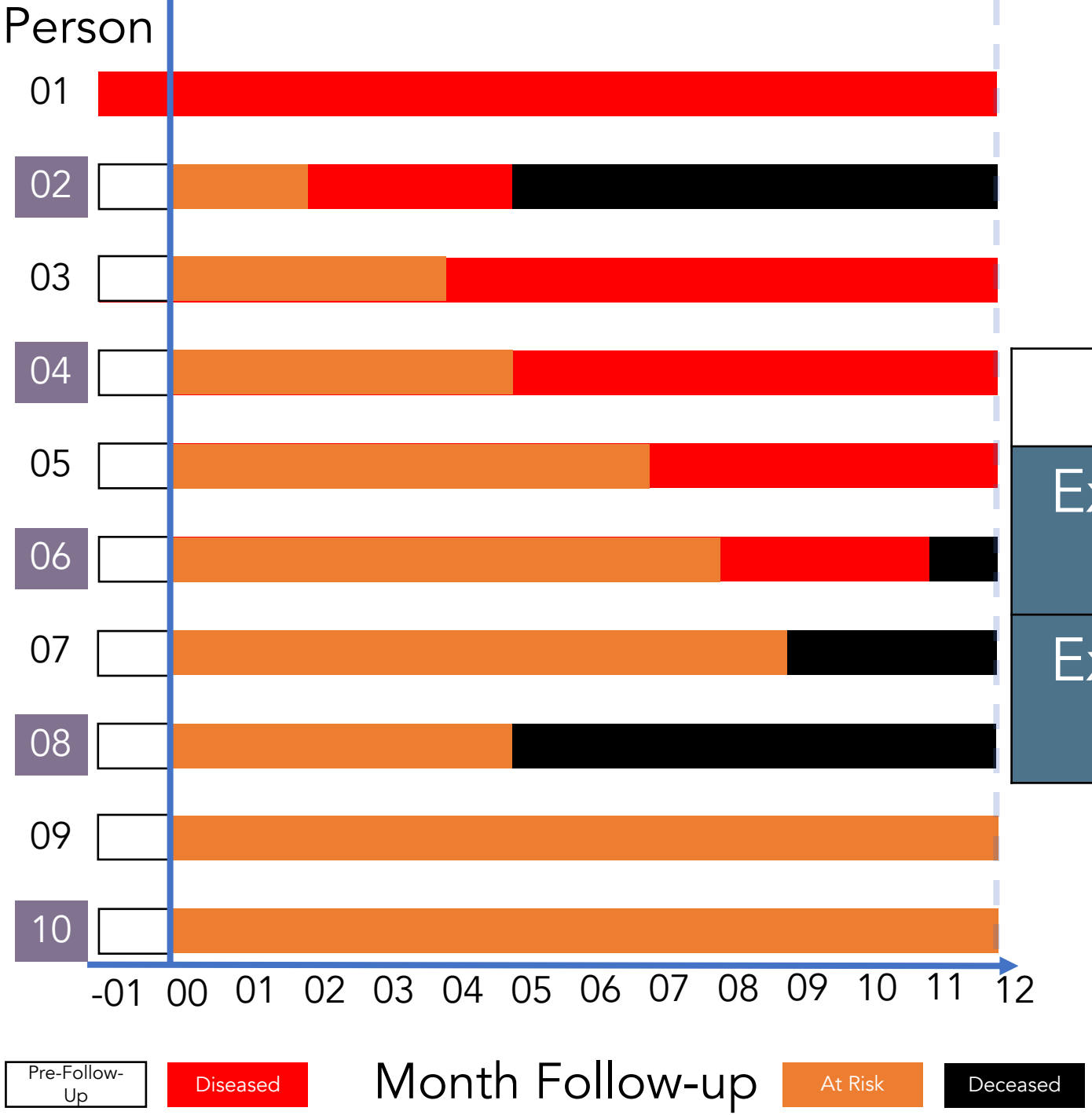
$$Incidence\ proportion\ in\ the\ exposed\ -Incidence\ proportion\ in\ the\ unexposed$$

Incidence Proportion Difference

	Outcome +	Outcome -	Row Total
Exposure +	a	b	a+b
Exposure -	c	d	c+d
Column Total	a+c	a+c	a+b+c+d

The incidence proportion among exposed = a / (a + b)
The incidence proportion among unexposed = c / (c + d)

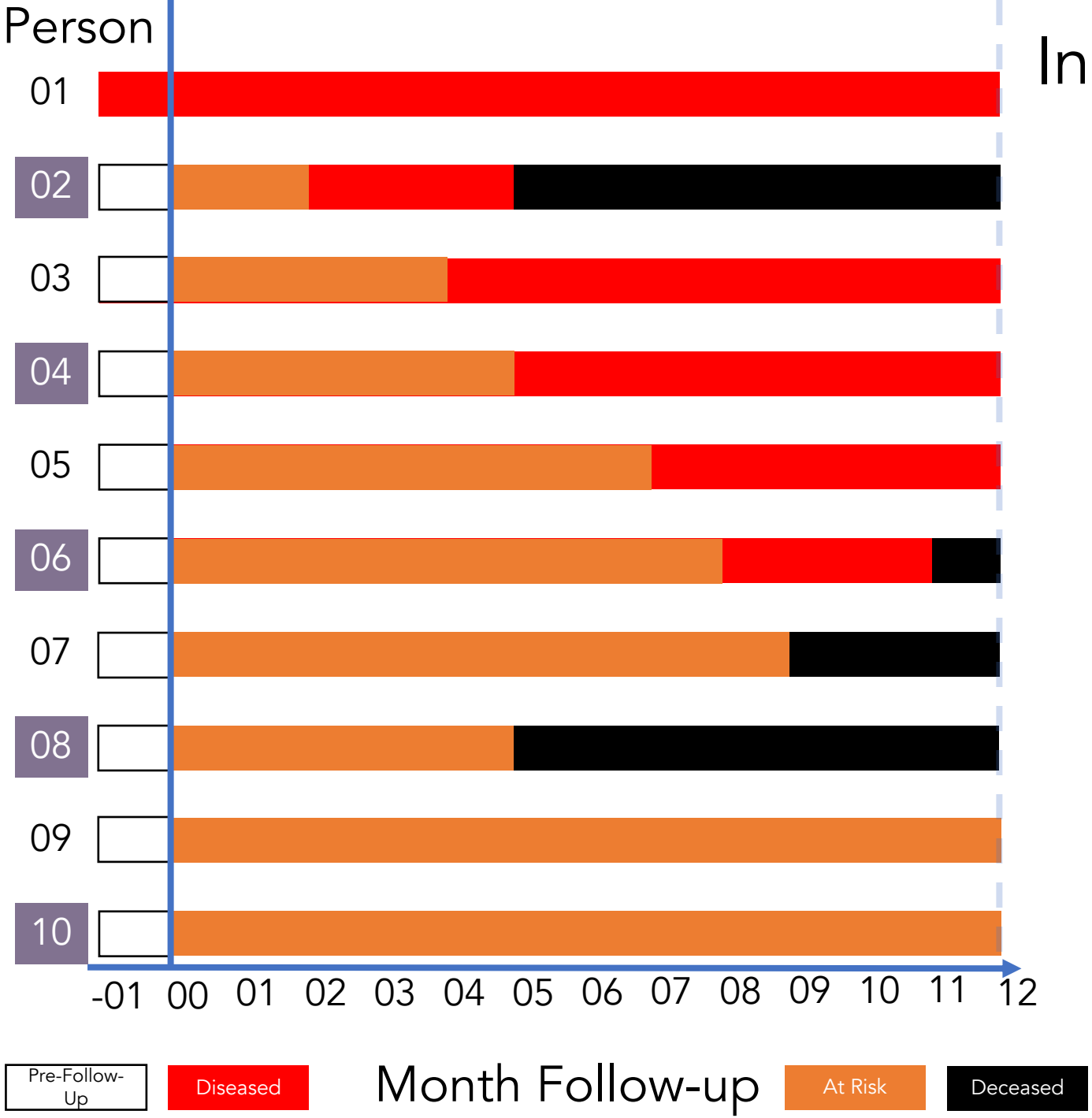
Incidence Proportion Difference $= \frac{a}{a+b} - \frac{c}{c+d}$



	Disease +	Disease -
Exposure +	3	2
Exposure -	2	2

Source: https://www.r-project.org/logo/

Incidence Proportion Difference



Among the members of our population, those who were exposed had 0.1 additional cases of disease per person compared to those who were not exposed over the 12 months of follow-up period.

Incidence Odds Ratio

- More commonly called
  - Odds ratio

Incidence Odds Ratio

$$\frac{Incidence\ Odds\ in\ the\ exposed}{Incidence\ Odds\ in\ the\ unexposed}$$

Incidence Odds Ratio

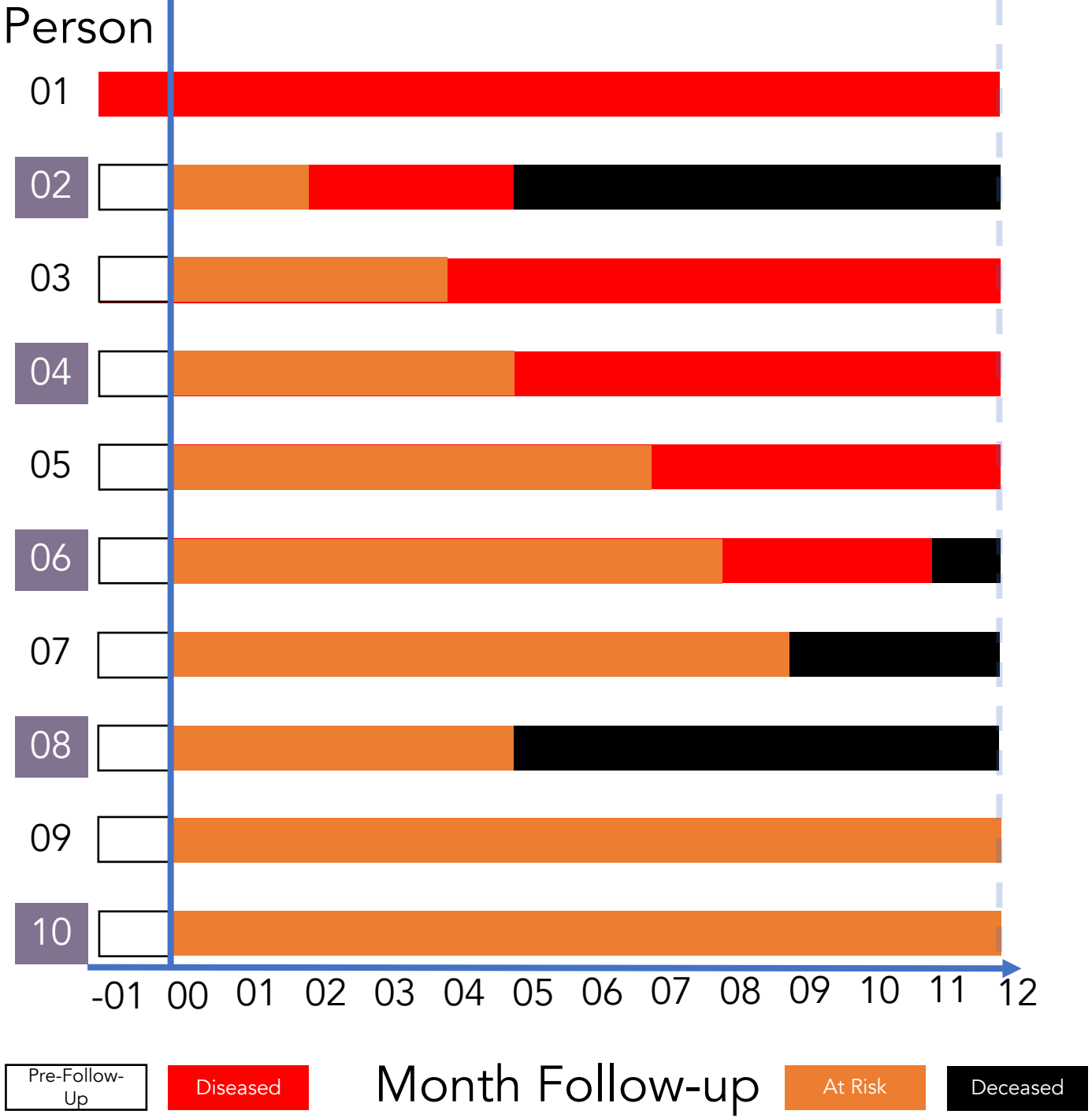
	Outcome +	Outcome -	Row Total
Exposure +	a	b	a+b
Exposure -	c	d	c+d
Column Total	a+c	a+c	a+b+c+d

The incidence odds among exposed = (a / (a + b)) / 1 - (a / (a + b))

The incidence odds among unexposed = (c / (c + d)) / 1 - (c / (c + d))

$$\text{Incidence Odds Ratio} = \frac{(a / (a + b)) / 1 - (a / (a + b))}{(c / (c + d)) / 1 - (c / (c + d))} = \frac{a/b}{c/d}$$

Source: https://www.r-project.org/logo/



Incidence Odds Ratio

Among the members of our population, those who were exposed had 1.5 times the odds of incident disease compared to those who were not exposed over the 12 months of follow-up period.