Observational Study Designs

BIOS 6611

CU Anschutz

Week 5

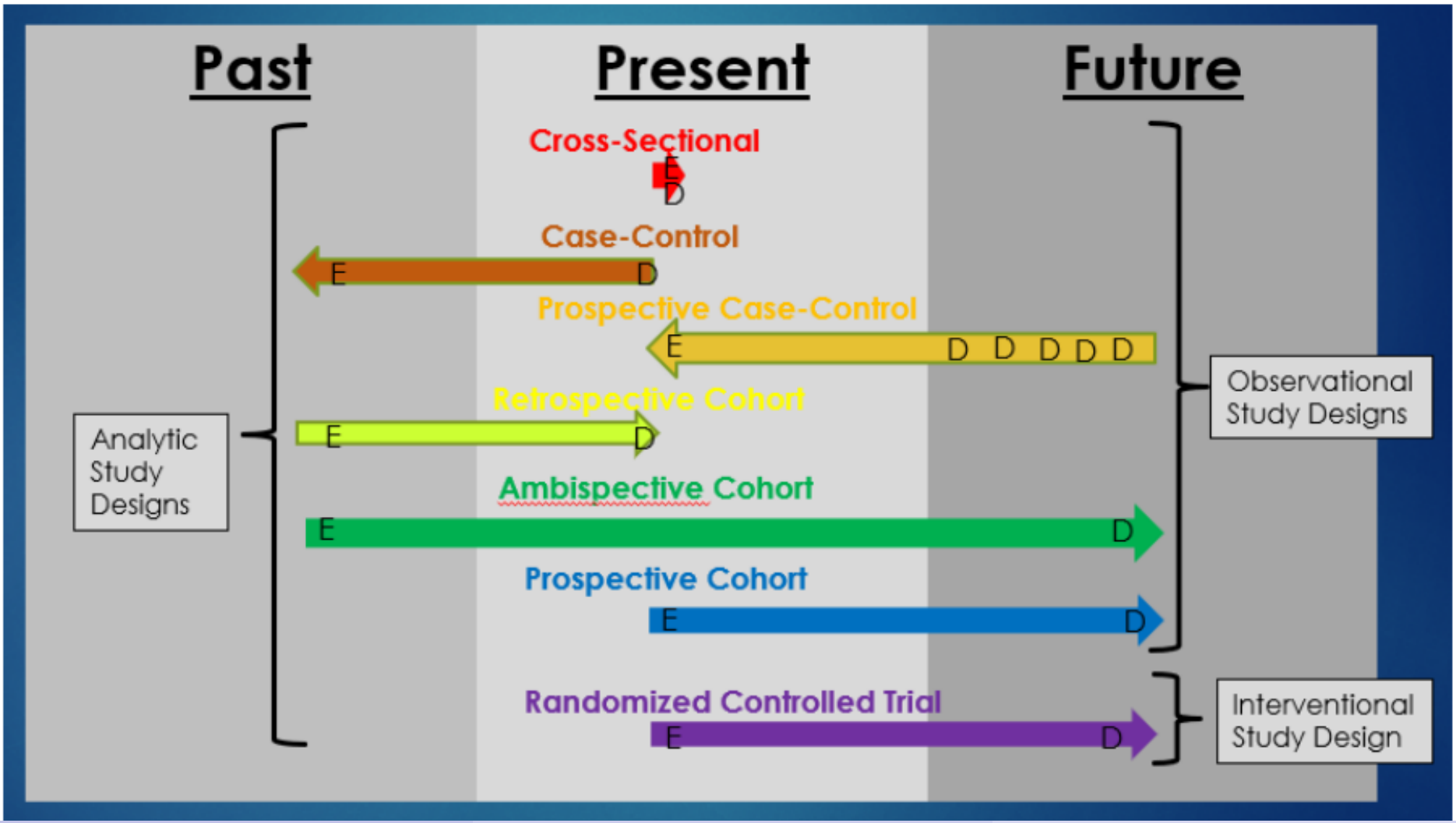
Observational Study Designs

Observational Studies

- **Observational studies**: researcher documents naturally occurring relationship between exposure and outcome. Researcher does not actively intervene. Exposure is already decided naturally or by some other factor.
  - Ex: looking at incidence of cavities in ice cream eaters and non ice cream eaters
- In contrast, **interventional studies** or **experiments** are when researcher actively performs an intervention
  - Ex: researcher randomly assigns one group to eat ice cream every day for 1 month, and the other group to abstain from eating ice cream. Researcher assesses the effect of ice cream eating on the risk of developing cavities.

Timing

Observational studies can be **prospective**, **retrospective**, or **cross sectional**.



Prospective Study

- A **prospective study** identifies a group of disease free individuals. The exposure is measured at baseline, then they are followed over some period of time until some develop the disease.
  - Study population is often referred to as a **cohort**. Sometimes will be called **prospective cohort studies.**
  - Ex: start with group with no cavities. Ask if regular ice cream eaters. Follow for 1 year, see how many develop cavities.
  - Less commonly, can also have **prospective case-control studies**, where controls are matched to cases as cases occur in real time.

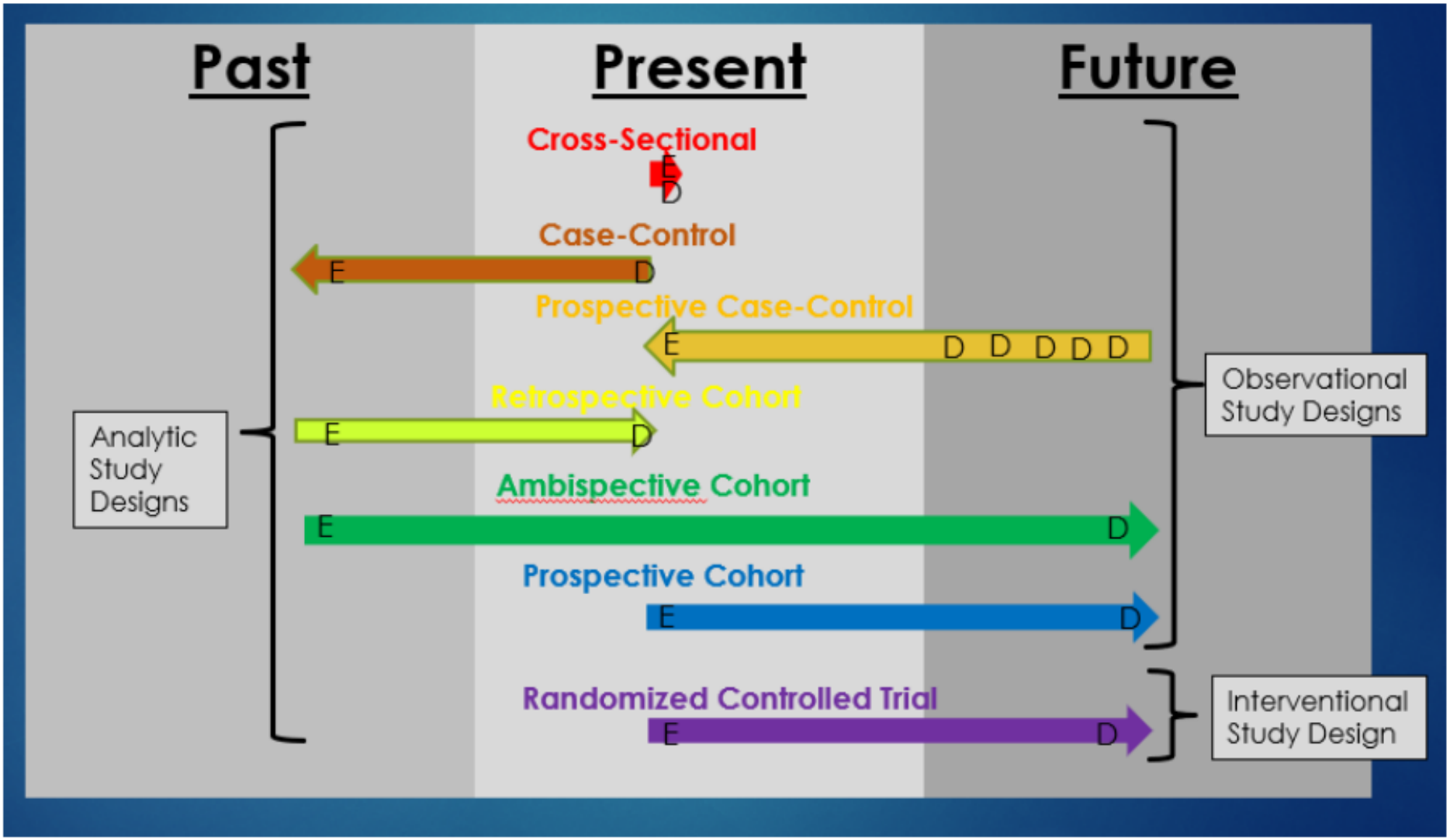
Retrospective Study

- A **retrospective study** initially identifies two groups: a group that has the disease (cases) and a group that does not (controls). Then looks at the relationship between prior exposures and disease status.
  - Also called a **case-control** study.
  - Ex: Find a group that has cavities (cases) and a group that does not have cavities (controls). Ask them about ice cream eating habits over past year.
  - Cases are often "matched" to controls that are similar (age, gender, etc.)
  - Less commonly, can also have **retrospective cohort studies**, where subjects are selected based on past exposure status and then researcher looks at the outcome's current status.

Cross-Sectional Study

- A **cross-sectional study** looks at the study population at a single point in time. All individuals are asked about current disease status and current exposure status.
  - Also called a **prevalence** study, because the prevalence of disease at one point in time is compared between exposed and unexposed individuals
  - Ex: ask group of people if currently have cavities and if are currently regular ice-cream eaters.

Observational Study Designs



Observational studies often result in 2 by 2 tables.

Intro to 2 by 2 Contingency Tables

Types of Random Variables

- Variables can be classified as **numeric** or **categorical**
- Numeric random variables can be classified as either **continuous** or **discrete**
  - Continuous random variables are quantitative values that can take on any value on a number line. They are usually obtained my measuring (ex: height, weight)
  - Discrete random variables can only taken on certain values. They are usually obtained by counting (ex: gene counts, successes)
- Categorical random variables can be classified as either **ordinal** or **nominal**
  - Ordinal random variables have ordered labels, but no quantitative value (ex: low, medium, high)
  - Nominal random variables have labels, but no inherent order (ex: dog breed)

Nominal Random Variables

- We will talk about measures for **nominal** random variables.
- Specifically, the situation with two possibly related nominal random variables, each with two categories.
- This kind of data is often represented in a 2×2 table. These are also called **contingency** tables.

Drinking Status	*Lung Cancer* Yes	No	
Heavy	33	1667	1700
Non	27	2273	2300
	60	3940	4000

2 by 2 Contigency Tables

In general, 2×2 contingency tables put the exposure (independent variable) on the left, and the disease (dependent variable) on the top. The row and column sums are also shown.

	Disease		
Exposure	Yes	No	
Yes	a	b	$a+b = n_1$
No	c	d	$c+d = n_2$
	$a+c = m_1$	$b+d = m_2$	

Measures of Effect for 2 by 2 Contingency Tables

Exposure	Disease: Yes	Disease: No	
Yes	a	b	$a+b = n_1$
No	c	d	$c+d = n_2$
	$a+c = m_1$	$b+d = m_2$	

We are interested in the probability of disease for exposed and unexposed subjects.

Let $p_1 = \frac{a}{n_1}$ be the estimate for the probability of disease among exposed.

Let $p_2 = \frac{c}{n_2}$ be the estimate for the probability of disease among unexposed.

Measures of Effect for 2 by 2 Contingency Tables

3 ways to describe behavior of p_1 and p_2:

1. Risk Difference (RD)
2. Risk Ratio (RR)
3. Odds Ratio (OR)

Note: p_1, p_2, RD, RR, and OR are statistics. They use the data to estimate the true probability parameters.