DECART Summer School 2018:

Causal Inference Module

Counterfactual Outcomes and Assumptions

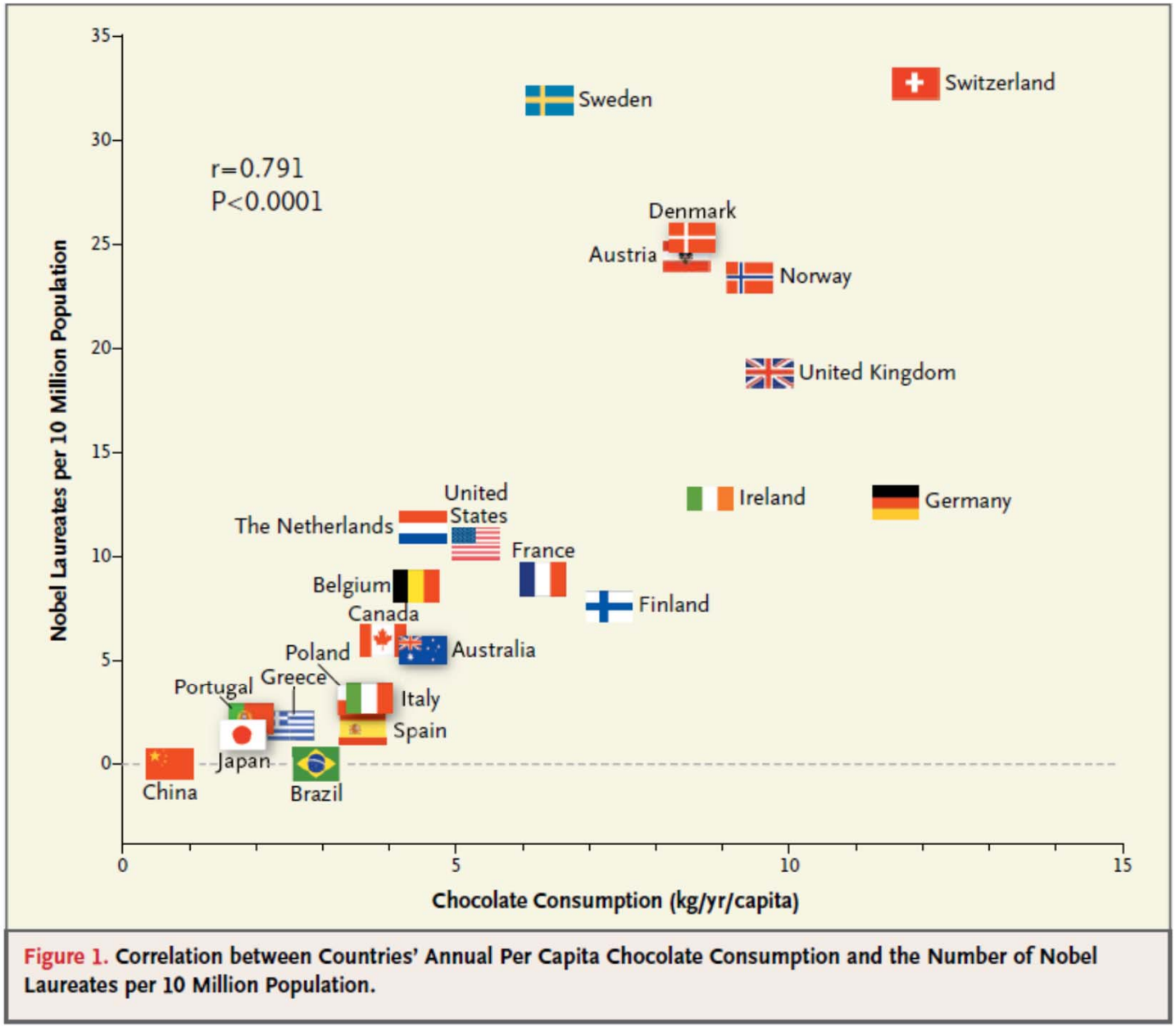
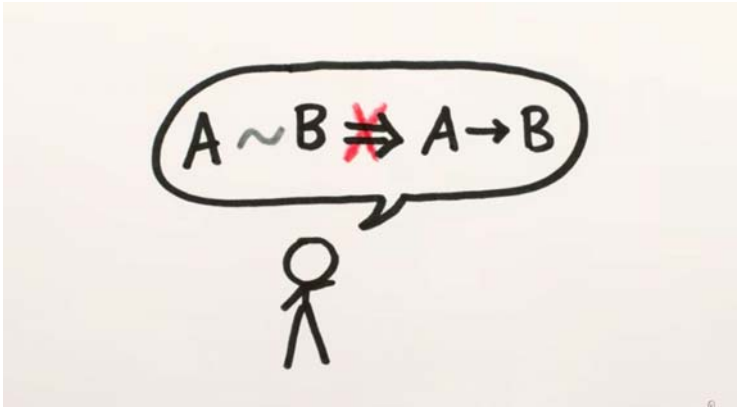


Figure 1. **Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.**

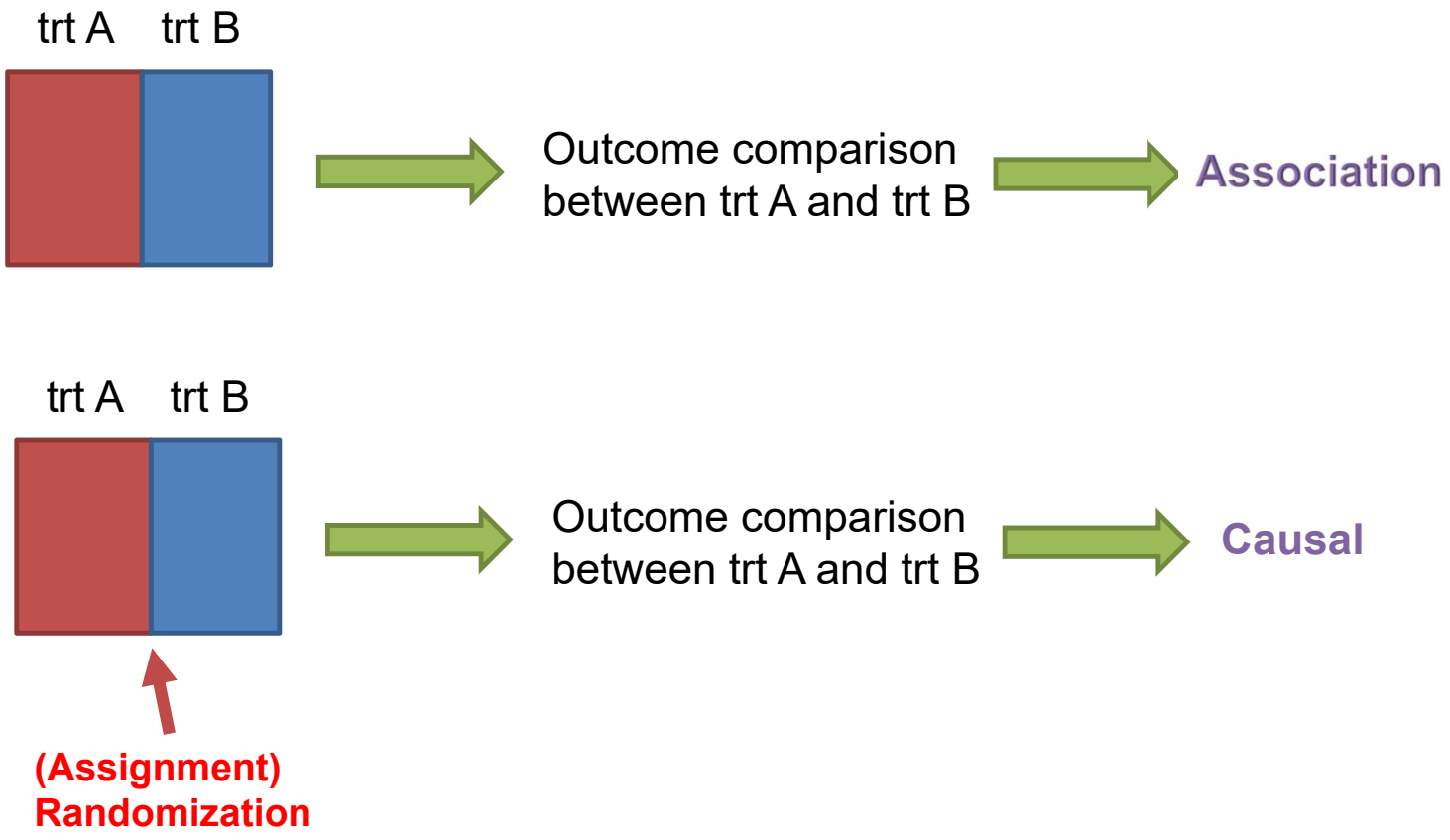
originally NEJM 2012

More can be found at http://www.tylervigen.com/spurious-correlations

Distinguish Causality



- X causes Y ⇒ X and Y are associated
- X causes Y ⇏ X and Y are associated



What Does Causality Mean?

Change the value of X

⇒ Change the distribution of Y

(change parameter or the distribution structure)

Counterfactual Framework of Modern Causal Inference

- Rubin Causal Model (RCM; Rubin 1978) is an approach to the statistical analysis of cause and effect based on the framework of potential outcomes (one of the most popular approaches in causal inference)
- J. Neyman first used the term 'potential outcome' in his Master's thesis (1923, Polish)

Neyman, Jerzy. Sur les applications de la theorie des probabilites aux experiences agricoles: Essai des principes. Master's Thesis (1923)

- In 1990, D. M. Dabrowska, and T. P. Speed translated it and reprinted on *Statistical Science* (Neyman-Rubin Model)

Neyman (1990 [from 1923]) On the application of probability theory to agricultural experiments. Essay on principles. Section 9. Statistical Science 5:465-480.

- D. Rubin (independently) brought this concept into a general framework for thinking about causation in both observational and experimental studies

D Rubin (1974) Estimating causal effects of treatments in randomized and nonrandomized studies. J Educational Psychology 66:688-701

- Accoridng to J. Heckman, potential outcomes model was first proposed in economics (Roy Model)

Roy, A. (1951): Some Thoughts on the Distribution of Earnings. Oxford Economic Papers 3(2), pp. 135-146.

- D. Rubin had a short comment on Heckman's citation in his Fisher Lecture "Causal Inference Using Potential Outcomes: Design, Modeling, Decisions" (Rubin 2005 JASA)

"In the economics literature, the use of the potential outcomes notation to define causal effects has recently (e.g., Heckman 1996) been attributed to Roy (1951) or Quandt (1958), which is puzzling because neither of these articles addresses causal inference, and the former has no mathematical notation at all. For seeds of potential outcomes in economics, the earlier references cited at the start of this paragraph are much more relevant; see the rejoinder by Angrist, Imbens, and Rubin (1996) for more on this topic."

Complete Counterfactual Framework for a Scalar Outcome and Single Treatment

- Study sample consists of patients (or subjects) indexed by i = 1, 2, …, n.
- Observe outcome Y_i for each subject i
- Observe treatment A_i for each subject i
- Also observe the method by which the treatment is administered (denoted K_i)
- Define the potential outcome $Y_i(a_1,k_1, a_2,k_2, \ldots, a_n,k_n)$ to be the outcome that *would be observed* for the i^{th} subject if each subject's treatment A_i and method of administration K_i were set to a_i and k_i, respectively, for i = 1, 2, …, n

Stable Unit Treatment Value Assumption (SUTVA)

SUTVA can be expressed as:

$$Y_i(a_1,k_1, a_2,k_2, \ldots, a_n,k_n) = Y_i(a_i)$$

irrespective of the values of k_j, $j = 1, 2, \ldots, n$, and irrespective of the values of a_j, $j \neq i$.

In other words, SUTVA asserts that the potential outcome for subject i is not affected by either the mode of delivery of the treatment or by the treatments received by the remaining subjects.

SUTVA has two components:

1. No interference (units do not interfere with each other): treatment applied to one unit does not effect the outcome for another unit
2. There is only a single version of each treatment level (potential outcomes must be well defined)

Big simplification: If SUTVA holds, we can express the potential outcome for subject i simply as $Y_i(a_i)$.

Consistency

- Closely related to SUTVA is the condition of consistency: $Y_i(a_i) = Y_i$ if $A_i = a_i$.
- Links counterfactual “universe” to observed “universe”
- Sometimes taken as an axiom if SUTVA is satisfied

Counterfactual Framework under SUTVA and Consistency

- Study sample consists of patients (or subjects) indexed by i = 1, 2, …, n.
- Observe outcome Y_i for each subject i
- Observe treatment A_i for each subject i
- The potential outcome $Y_i(a)$ is the outcome that *would be observed* for the i^{th} subject if this subject's treatment were set to a
- $Y_i(a') - Y_i(a)$ = Individual causal effect of changing a to a' for the i^{th} patient
- Alternative expressions for individual causal effects are possible, such as $Y_i(a')/Y_i(a)$. But unless indicated otherwise, we write the individual causal effect as $Y_i(a') - Y_i(a)$.
- $Y_i = Y_i(A_i)$ is the value of $Y_i(a)$ that we are able to observe

Counterfactual Framework Under SUTVA and Consistency

- For simple comparisons of outcomes with vs. without a single treatment, we let $A_i = 1$ if the treatment is given and $A_i = 0$ otherwise, and say the causal effect for the i^{th} patient is $Y_i(1) - Y_i(0)$.
- For such 2-valued treatments, the average causal effect in the study sample is: $\frac{1}{n}\sum_{i=1}^{n}[Y_i(1) - Y_i(0)]$ (ATE)
- Restricting this average to subjects with $A_i = 1$ provides the average causal affect in the treated (ATT)
- Restricting this average to subjects with $A_i = 0$ provides the average causal affect in the untreated (ATU)
- Obvious extension to multivalued treatments:

$$\frac{1}{n}\sum_{i=1}^{n}[Y_i(a') - Y_i(a)]$$

- When expressed in the full population, we can write:

$$ATE = E[Y_i(1) - Y_i(0)]$$

$$ATT = E[Y_i(1) - Y_i(0)|A_i=1]$$

$$ATU = E[Y_i(1) - Y_i(0)|A_i=0]$$

Why Randomization Works

- Randomization assures that A_i is statistically independent of all subject characteristics
- This includes $Y_i(a)$ for all a
- We write this as $A_i \perp Y_i(a)$ for all a
- The $Y_i(a)$ that we get to observe are an unbiased random sample of all of the $Y_i(a)$
- Thus, $E[Y_i|A_i=a] = E[Y_i(a)|A_i=a]$ (consistency)

  $= E[Y_i(a)]$ (randomness)
- For 2-valued treatments, we thus have

$$E[Y_i|A_i=1] - E[Y_i|A_i=0] = E[Y_i(1)] - E[Y_i(0)] = E[Y_i(1) - Y_i(0)]$$

Conditional Randomization

- Let L_i designate a discrete random variable, possibly related to $Y_i(0)$ and $Y_i(1)$. L_i could be a function of other patient characteristics; e.g.,

  $L_i = 1$ if age < 30 and female

  $L_i = 2$ if age < 30 and male

  $L_i = 3$ if age ≥ 30 and female

  $L_i = 4$ = age ≥ 30 and male

- Suppose A_i is randomized within each stratum defined by L_i, with $Pr(A_i=1)=e_i(L_i) = e(l)$, say, where l is the value taken by L_i
- Thus $A_i \perp Y_i(0) \mid L_i$ and $A_i \perp Y_i(1) \mid L_i$

  (A_i is conditionally independent of $Y_i(0)$ and $Y_i(1)$ given L_i)
- By consistency and conditional independence,

  $E[Y_i|A_i=1, L_i = 1] = E[Y_i(1)|A_i=1, L_i = 1] = E[Y_i(1)| L_i = 1]$ and

  $E[Y_i|A_i=0, L_i = 1] = E[Y_i(0)|A_i=0, L_i = 1] = E[Y_i(0)| L_i = 1]$.
- Let $\overline{Y}_{1l}$ and $\overline{Y}_{0l}$ denote sample means for treated and untreated in stratum l
- We have $E(\overline{Y}_{1l}) = E[Y_i(1)|L_i= l]$ and $E(\overline{Y}_{0l}) = E[Y_i(0)|L_i= l]$

Observational Studies

- Treatment assignment A is (almost?) never truly random
- In fact, treatment assignment A is (almost?) never truly conditionally random
- But, perhaps in some cases, we can find covariates X_1, X_2, …, X_p such that the treatment assignment is unrelated to baseline factors that predict the outcome for fixed X_1, X_2, …, X_p

Potential Outcomes and Causal Effects

- Average Causal Effect

$$\theta = E\left(Y_i^1\right) - E\left(Y_i^0\right)$$

- Association

$$\alpha = E(Y_i|A_i = 1) - E(Y_i|A_i = 0)$$

- Association is not causation, i.e. $\theta \neq \alpha$ in general

Example of Potential Outcomes and Causal Effects

index	A_i	Y_i	$Y_i(0)$	$Y_i(1)$	$Y_i(1) - Y_i(0)$
1	0	4	4	3	-1
2	0	7	7	5	-2
3	0	2	2	7	5
4	0	8	8	9	1
5	1	3	2	3	1
6	1	5	4	5	1
7	1	8	5	8	3
8	1	9	3	9	6

$$\hat{\theta} = \frac{1}{8} \times (3 + 5 + 7 + 9 + 3 + 5 + 8 + 9) - \frac{1}{8} \times (4 + 7 + 2 + 8 + 2 + 4 + 5 + 3) = 1.75$$

$$\hat{\alpha} = \frac{1}{4} \times (3 + 5 + 8 + 9) - \frac{1}{4} \times (4 + 7 + 2 + 8) = 1$$

Potential Outcomes and Causal Effects

- Average Causal Effect

$$\theta = E\left(Y_i^1\right) - E\left(Y_i^0\right)$$

- Association

$$\alpha = E(Y_i|A_i = 1) - E(Y_i|A_i = 0)$$

- Association is not causation, i.e. $\theta \neq \alpha$ in general
- Can we estimate causal effect?

  Yes! ⇒ need additional assumptions

Standard Assumptions of Causal Inference

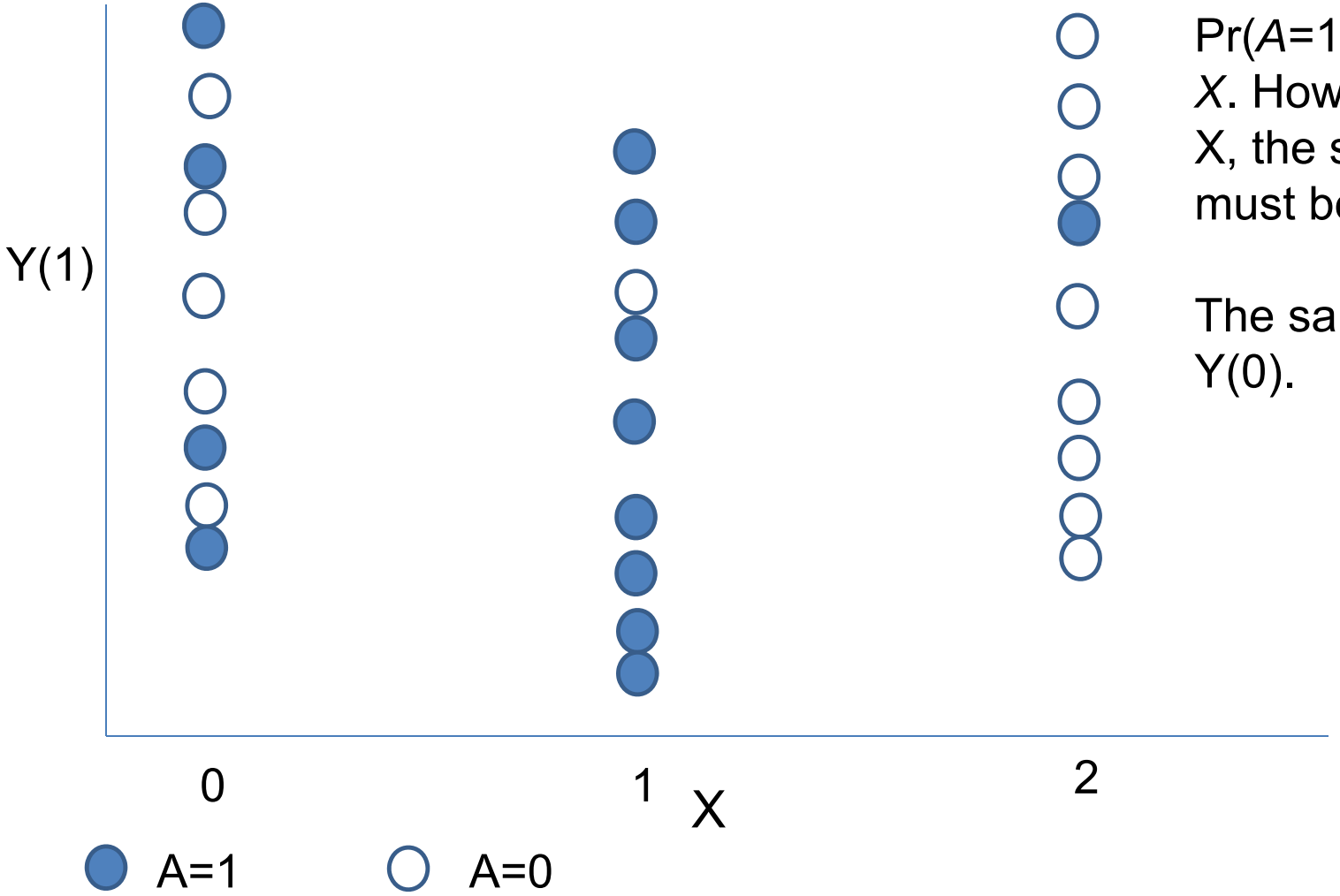
1. ***Conditional Exchangeability (No Unmeasured Confounding Assumption, NUCA)*:**

A is statistically independent of Y(1) and Y(0) conditional on $X_1, X_2, \ldots, X_p$.

- Conditional independence is often written as

  $A \perp Y(1)$ and $A \perp Y(0)$ given $X_1, X_2, \ldots, X_p$.
- Intuitively, conditional exchangeability implies that the p covariates included in the analysis are sufficient to control confounding
- Conditional randomization would imply that A is conditionally independent of ***all*** other subject characteristics given $X_1, X_2, \ldots, X_p$. For causal inferences concerning the effects of A on Y, it is sufficient that A is conditionally independent of Y(1) and Y(0) given $X_1, X_2, \ldots, X_p$.
- In other words, conditional randomization is a stronger condition than conditional exchangeability. But conditional exchangeability is enough for the purposes of causal inference.

Illustration of Conditional Exchangeability



Pr(*A*=1) can depend on *X*. However, for any given X, the subjects with A=1 must be unrelated to Y(1).

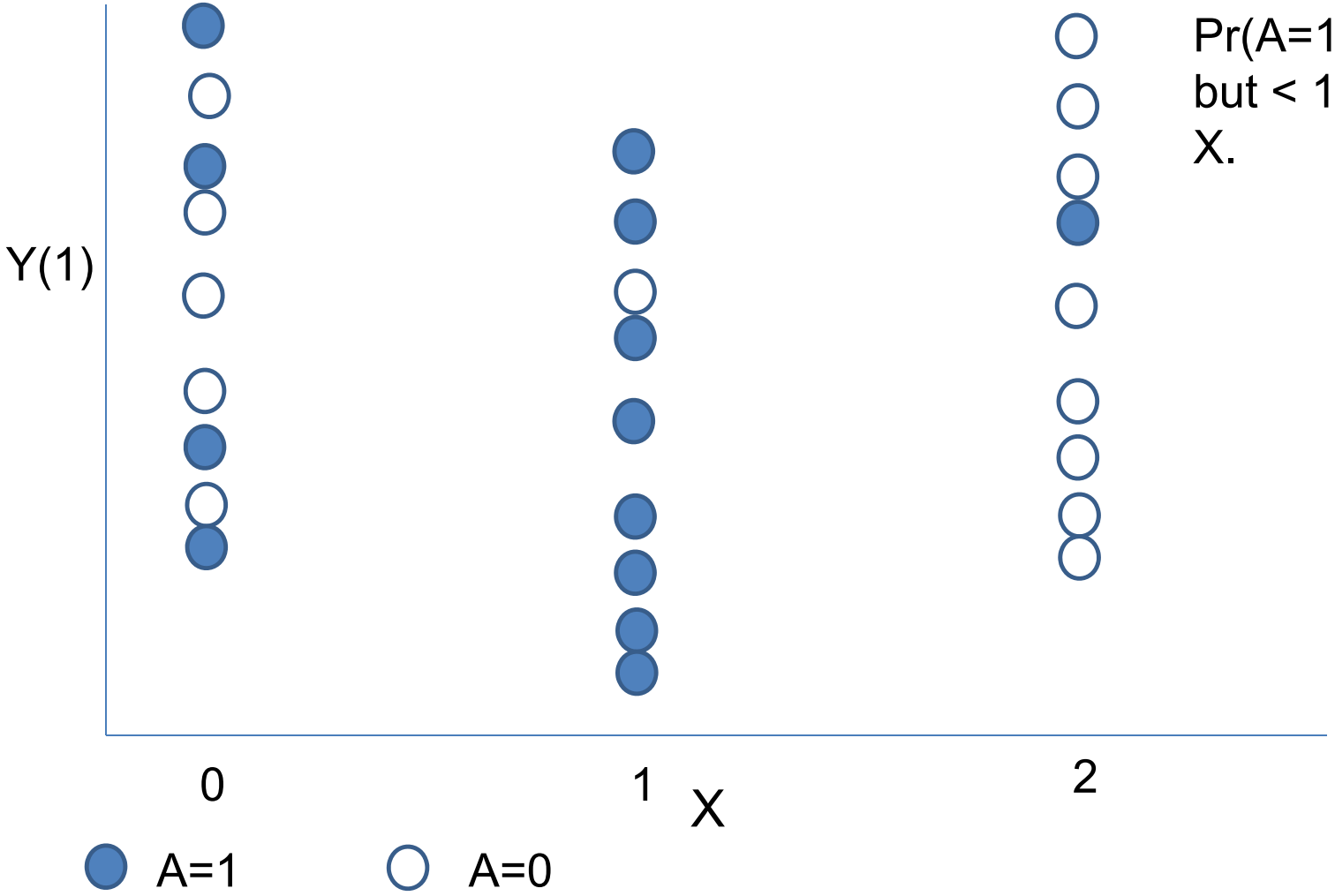
The same most hold for Y(0).

Standard Assumptions of Causal Inference

2. ***Positivity***: $0 < \Pr(A = 1 | X_1, X_2, \ldots, X_p) < 1$ for all values of $X_1, X_2, \ldots, X_p$ that occur in the study population.

- Positivity requires that all possible values of the treatment (treated and untreated for the examples we have been considering) may occur for every combination of values for the covariates that occur in the study population
- Positivity can be thought of as a type of equipoise condition.
- Positivity is not absolutely required when performing causal inference using regression analysis. However, if positivity is violated, the analysis will require extrapolation of the regression model beyond the range of the data, which is very risky and usually is not appropriate.

Illustration of Positivity Assumption



Pr(A=1) must be > 0 but < 1 for each value of X.