Week 7 – Thursday Session

Standardization

EPI202 – Epidemiologic Methods II
Murray A. Mittleman, MD, DrPH
Department of Epidemiology, Harvard TH Chan School of Public Health

HARVARD T.H. CHAN
SCHOOL OF PUBLIC HEALTH

Standardization with Conditional Probabilities
Standardized Observed Cumulative Incidence Pr[Y=1|A=a]

A=exposure
Y=outcome
L=third variable

- The standardized observed cumulative incidence Pr[Y=1|A=a] is the weighted average of the stratum specific cumulative incidences
  - Pr[Y=1|L=0, A=a] and
  - Pr[Y=1|L=1, A=a]

  with weights equal to the proportion of exposed or unexposed people in the population with L=0 and L=1

- **$\Pr[Y=1|A=a]=\Sigma_\ell \Pr[Y=1|L=\ell,A=a]\Pr[L=\ell]$**
  - where Σ_ℓ means sum over all values ℓ that occur in the population
  - e.g. Pr[Y=1|A=1]=Pr[Y=1|L=0,A=1]Pr[L=0]+Pr[Y=1|L=1,A=1]Pr[L=1]
  - e.g. Pr[Y=1|A=0]=Pr[Y=1|L=0,A=0]Pr[L=0]+Pr[Y=1|L=1,A=0]Pr[L=1]

Hernán MA, Robins JM. Causal Inference.

Standardization with Conditional Probabilities
Standardized Counterfactual Cumulative Incidence $Pr[Y^a=1]$

A=exposure
Y=outcome
L=third variable

- The standardized counterfactual cumulative incidence $Pr[Y^a=1]$ is the weighted average of the stratum specific cumulative incidences
  - $Pr[Y^a=1|L=0]$ and
  - $Pr[Y^a=1|L=1]$

  with weights equal to the proportion of people in the population with L=0 and L=1

- $\mathbf{Pr[Y^a=1]=\Sigma_{\ell}Pr[Y^a=1|L=\ell]Pr[L=\ell]}$
  - where Σ_{ℓ} means sum over all values ℓ that occur in the population
  - e.g. $Pr[Y^a=1]=Pr[Y^a=1|L=0]Pr[L=0]+Pr[Y^a=1|L=1]Pr[L=1]$

- With consistency and positivity, $Pr[Y=1|A=a] = Pr[Y^a=1]$

Hernán MA, Robins JM. Causal Inference.

Inverse Probability Weighting (IPW)
Setting

- IPW can be used to adjust for measured confounding and selection bias under the assumptions of
  - consistency
  - exchangeability
  - positivity
  - no misspecification of the model used to estimate weights (in non-parametric setting)

- Similar to other analytic approaches, data is necessary on all sources of confounding and other sources of bias in order to account for it in the analysis

Conditional Exchangeability

- The investigators may believe that the exposed and the unexposed are exchangeable within levels of some variables L
  - Had exposed patients in critical condition stayed unexposed, they would have had the same mortality risk as those in critical condition who actually stayed unexposed (and vice versa)
  - And similarly for patients in noncritical condition

- That is, the investigators may be willing to assume conditional exchangeability
  - Often many factors must be taken into account
  - E.g., treated and untreated are not comparable because different risk of outcome by symptoms, access to care, SES, drug user etc.

Exchangeability Within Levels of the Stratification Factor(s)

- Consider only individuals with the same pre-exposure prognostic factors

- Then the exposed and the unexposed may be exchangeable
  - e.g., among individuals with an ejection fraction of 40%, those who do and do not receive a heart transplant may be comparable

- Sometimes reasonable, especially if conditioning on many pre-exposure covariates L

In the pseudo-population created with inverse probability weighting (IPW), exposure is independent of the:

- Outcome
- Unmeasured confounders
- Measured confounders
- Do not know

Inverse Probability Weighting (IPW)
How it Works

- IPW creates a pseudo-population in which the exposure is independent of the measured confounders

- The pseudo-population is the result of assigning a weight to each participant that is, informally, proportional to the participant's probability of receiving her own exposure history.

Heart Transplant and Death
Summarized in a Table

	Non-Critical Condition (L=0)			Critical Condition (L=1)		
	Heart Transplant (A=1)	No Heart Transplant (A=0)	Total	Heart Transplant (A=1)	No Heart Transplant (A=0)	Total
Death (Y=1)	1	1	2	6	2	8
Survival (Y=0)	3	3	6	3	1	4
Total	4	4	8	9	3	12

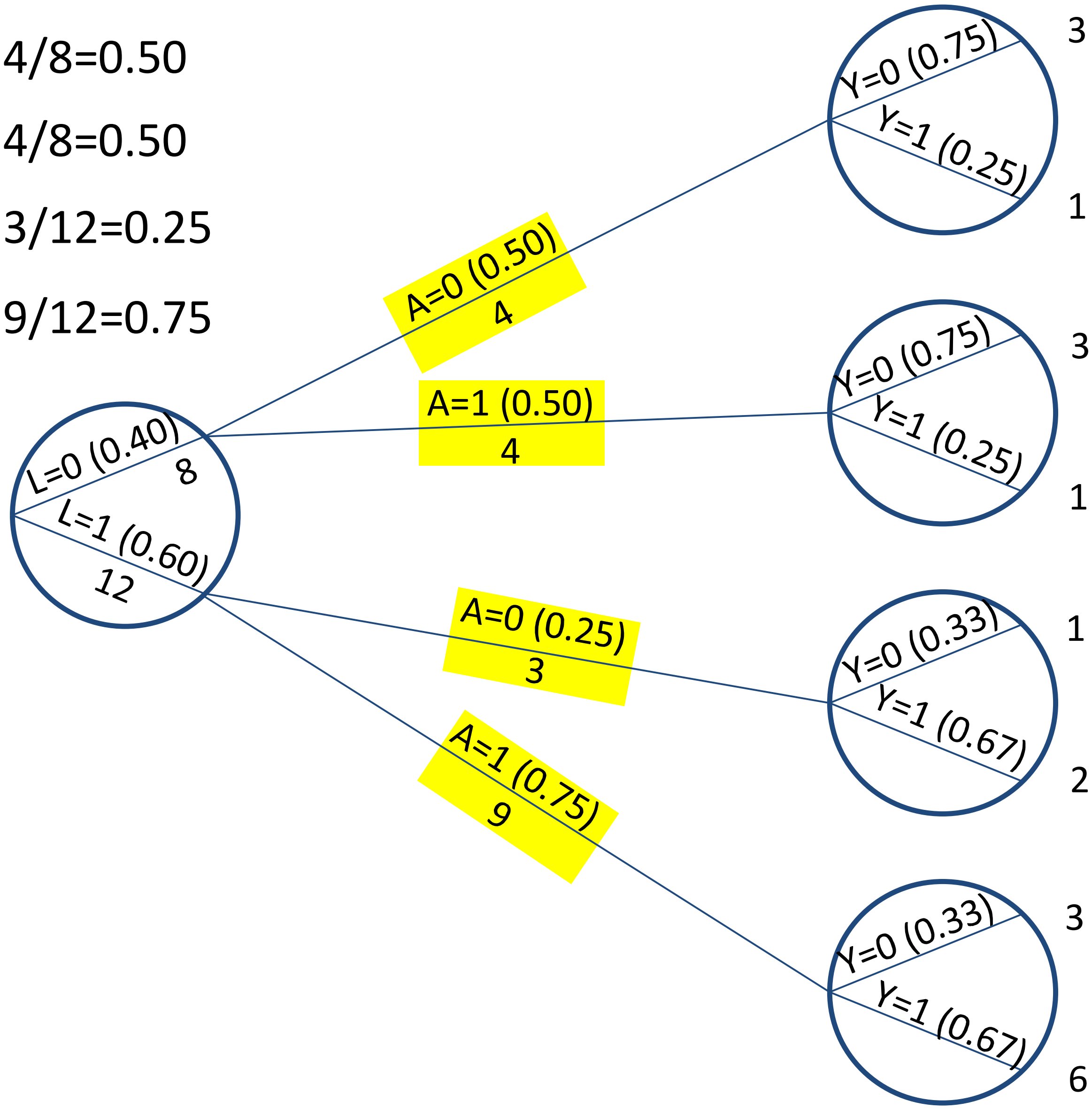
Heart Transplant and Death
Summarized in a Tree

Pr[A=0|L=0] = 4/8=0.50

Pr[A=1|L=0] = 4/8=0.50

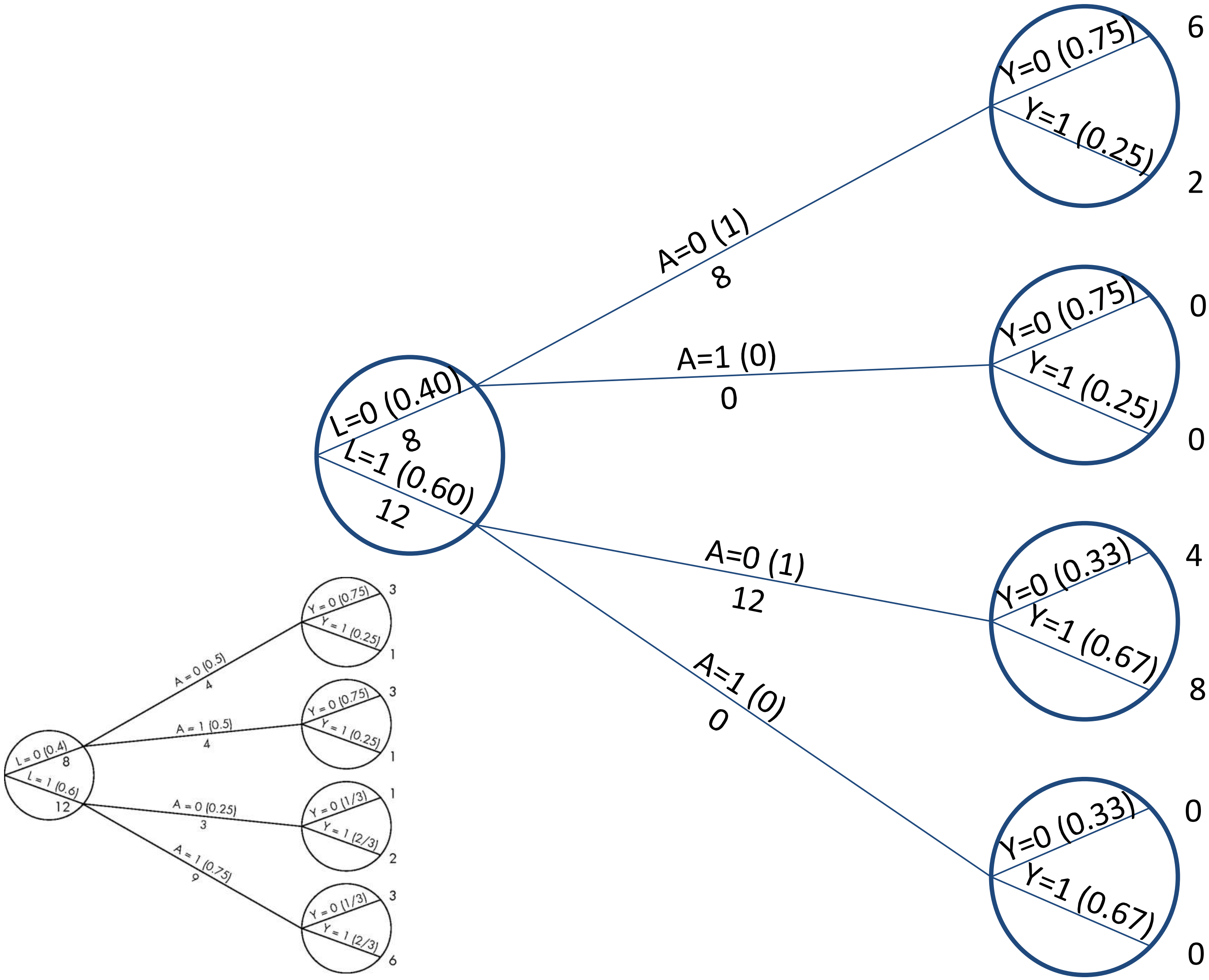
Pr[A=0|L=1] = 3/12=0.25

Pr[A=1|L=1] = 9/12=0.75



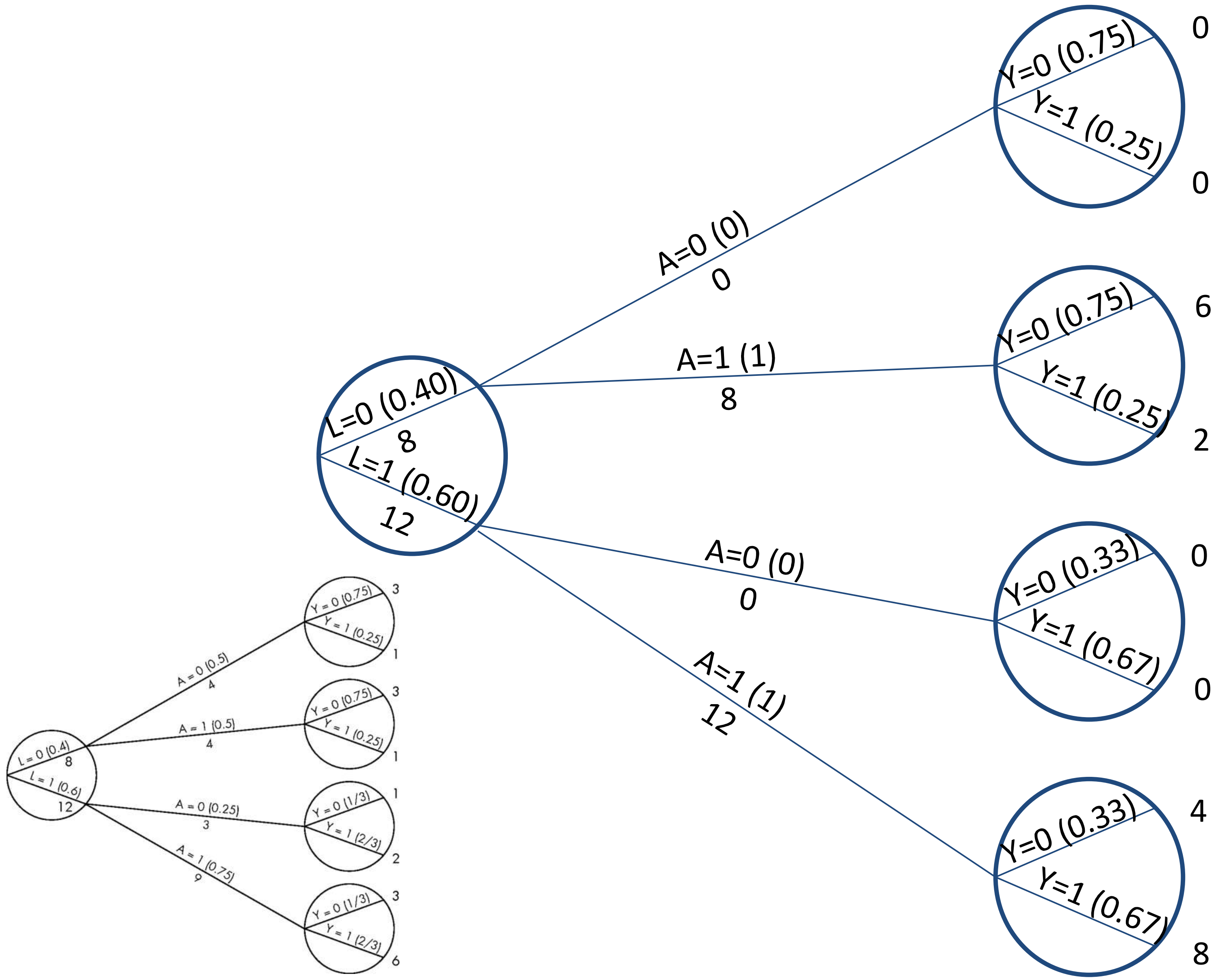
Heart Transplant and Death

Pr[$Y^{a=0}$=1]



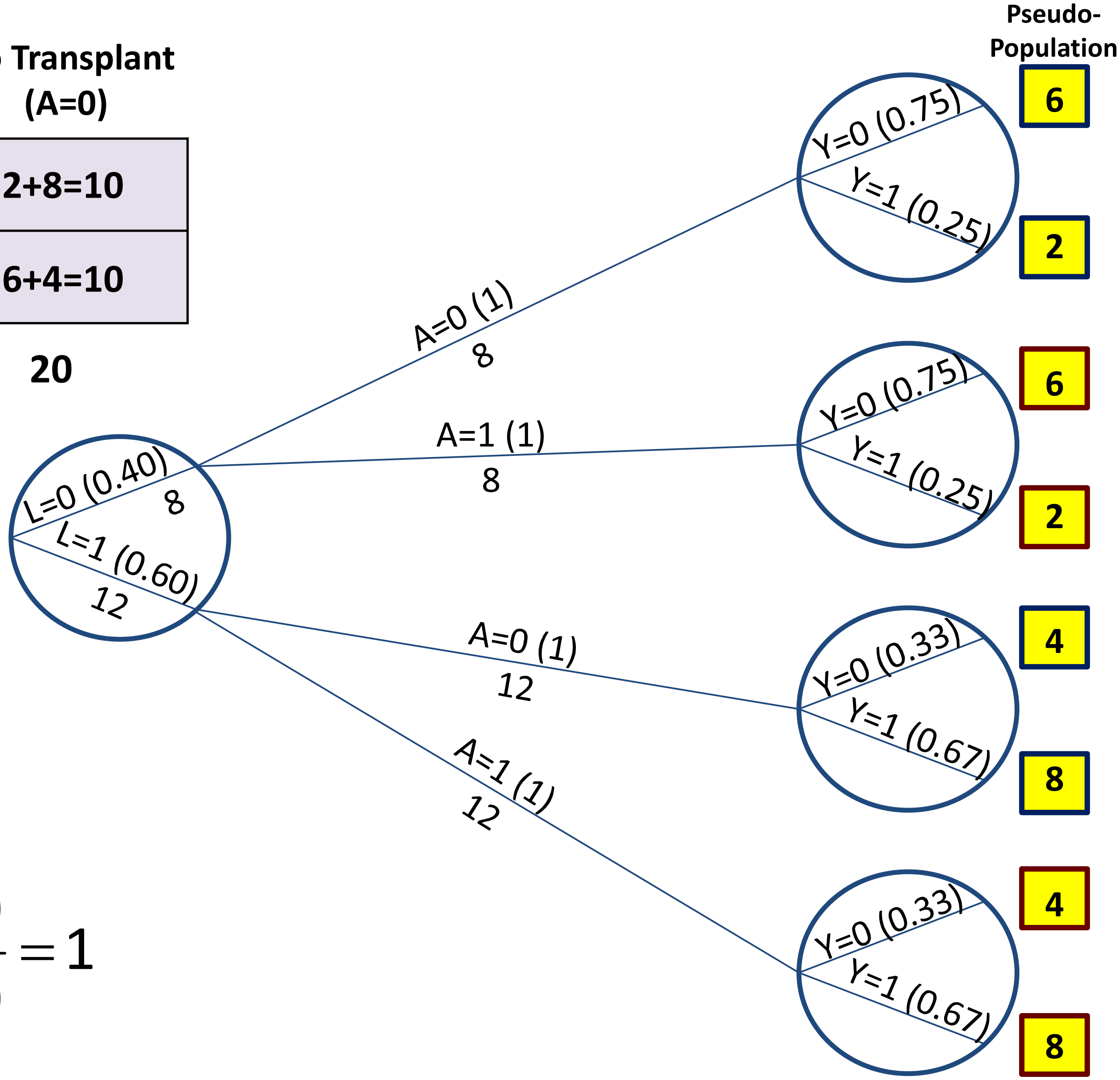
Heart Transplant and Death

$\Pr[Y^{a=1}=1]$



Heart Transplant and Death
Average Association for Total Population

	Transplant (A=1)	No Transplant (A=0)
Death (Y=1)	2+8=10	2+8=10
Survival (Y=0)	6+4=10	6+4=10
Total	**20**	**20**



$$\frac{\Pr[Y^{a=1}=1]}{\Pr[Y^{a=0}=1]} = \frac{10/20}{10/20} = 1$$

IPW Pseudopopulation

- When the required assumptions for causal inference are met, the pseudo-population created by IP weighting looks like a marginally randomized experiment
  - Exposed and unexposed subjects are (unconditionally) exchangeable
  - The pseudopopulation simulates what would have happen to each individual under both counterfactual scenarios
  - It is as if exposure is randomized, since exposure is equally probable across levels of the covariate L
  - There is no confounding

- In the pseudo-population, no adjustments are necessary to compute the causal effect

IPW=Standardization

- All approaches to standardization (unified, traditional and IPW) involve applying weights to estimate the measure of association that would be observed in a population with a particular distribution of stratification factor(s).

- For causal inference, all assume conditional exchangeability
  - $Y^a \coprod A|L=l$ for all a
  - no unmeasured confounding within levels of the measured variable(s) L

Weights for Standardization and IPW

- Standardization and IPW calculate different components of the joint distribution
  - IPW: f[A|L]
  - Standardization: f[L], f[Y|A,L]

A=exposure
Y=outcome
L=stratification factor

- The two methods yield results that are algebraically equivalent

	Standardization	**IPW Weight**
Total Population	$\frac{\Pr(Y^{a=1}=1)}{\Pr(Y^{a=0}=1)}$	$\frac{1}{f(A \mid L)}$
Exposed	$\frac{\Pr(Y^{a=1}=1 \mid A=1)}{\Pr(Y^{a=0}=1 \mid A=1)}$	$\frac{\Pr(A=1 \mid L)}{f(A \mid L)}$
Unexposed	$\frac{\Pr(Y^{a=1}=1 \mid A=0)}{\Pr(Y^{a=0}=1 \mid A=0)}$	$\frac{\Pr(A=0 \mid L)}{f(A \mid L)}$

Comparison of Standardization Approaches

- All standardization approaches are mathematically equivalent
  - Traditional (ratio or difference of standardized measures of occurrence)
  - Unified (weighted average of stratum specific measures of association)
  - Inverse probability weighting
- However, IPW calculates different component of the joint distribution, and does not require stratum-specific incidence rates
- So far in our class, all of the analyses are based on binary exposures and binary outcomes.
- Unlike traditional and unified approaches, IPW can be applied in a regression framework, allowing for adjustment of polytomous and continuous confounders and modifiers

Comparison of Summary Measures

- Recall that if there is no evidence of effect measure modification, the expected value of all summary measures are equivalent
  - Unified/Traditional standardization
  - IPW standardization
  - Mantel-Haenszel weighted average

- The MH weights will be more statistically efficient than standardization and may be preferred. However, if the assumption regarding freedom from effect measure modification is violated, the MH estimator (and other information weighted estimates, such as from conventional regression models) are not clearly interpretable.

- Logistic regression can also be applied to allow for adjustment of polytomous and continuous confounders and modifiers

- Standardization, including IPW has several theoretical advantages.
  - There is an explicit connection to the counterfactual contrasts of interest for causal inference
  - The standardized estimates are valid and clearly interpretable in the presence of effect measure modification

Assuming conditional exchangeability, positivity, consistency, a well-defined intervention and no effect measure modification, the Mantel-Haenszel estimator estimates the same (causal) parameter as the IPW estimate.

- True
- False
- Do not know

HAVE A GOOD WEEKEND