Week 7: Overview of Standardization

Video 1: Utility of Standardization

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

HARVARD T.H. CHAN
SCHOOL OF PUBLIC HEALTH

Key Concepts

- Typical weighted averages vs. standardization

- Standardization weights that reflect the distribution of the stratification factor(s) in the
  - Unexposed (direct standardization)
  - Exposed (indirect standardization)
  - Total study population

- Equivalent approaches to standardization
  - Weighted average of stratum-specific rate ratios or differences (unified)
  - Ratios or differences of weighted average of stratum-specific rates (traditional)
  - Inverse probability of treatment weighting (IPW)

So Far…
Exchangeability

- **Crude** assumes no confounding and no effect measure modification

- **If non-exchangeability**
  - Crude estimate is confounded
  - To attain exchangeability, stratify by levels of confounding factor(s) so that within strata, the exposed and unexposed are exchangeable
  - Assuming no residual confounding or other forms of bias, these stratified estimates are unbiased

So Far…

Typical Efficient Weighted Average Estimates

In the absence of true effect measure modification

- Assume all stratum-specific estimates arose from underlying distributions centered at the same true value of the population parameter

- Maximize efficiency by averaging across strata using MH weights (ratio measures) or MH-style weights (difference measures)

- Weights are driven by the data

So Far…
Effect Measure Modification

In the presence of true underlying effect measure modification

- Data-driven weights are inappropriate since summary measure does not reflect any particular population
  - e.g. beta-blockers and mortality

- Show stratum-specific estimates or… use standardization

- Standardization is an approach to creating weighted averages that reflect the distribution of stratification factor(s) in populations of interest

- Standardization does not assume homogeneity across strata

Aluminum Work and Mortality

Crude

	Al	No Al	
Deaths	54	450	504
Person-years	11,000	300,000	311,000
Incidence Rate	0.005	0.002	

IRR=3.27

$$\begin{tikzcd} \text{Age} \arrow[r] \arrow[rr, bend left] & \text{Al work} & \text{Death} \end{tikzcd}$$

Stratified by Age

Age	Young			Old		
	Al	No Al		Al	No Al	
Deaths	50	50	100	4	400	404
Person-years	10,000	100,000	110,000	1,000	200,000	201,000
Incidence Rate	0.005	0.0005		0.004	0.002	

IRR=10

IRR=2

Age-Specific Death Rates

	Aluminum Reduction Plant Workers				Regional Sample of Non-Plant Workers				IRR_i
	Person-Years	Deaths	Rate	%	Person-Years	Deaths	Rate	%	
	Person-Years	Cases	Rate	Weight	Person-Years	Cases	Rate	Weight	
Young	10,000	50	0.005	0.91	100,000	50	0.0005	0.33	10
Old	1,000	4	0.004	0.09	200,000	400	0.002	0.67	2
Total	11,000	54	0.004909		300,000	450	0.0015		

- $$IRR_{crude} = \frac{54}{11{,}000} \Big/ \frac{450}{300{,}000} = 3.27$$

$$\begin{tikzcd} \textbf{Age} \arrow[r] \arrow[rr, bend left] & \textbf{Al work} & \textbf{Death} \end{tikzcd}$$

- Age is a confounder: younger people
  - Are more likely to be exposed than older people: 91% of the person-time among workers were young whereas only 33% of the person-time among the regional sample were young.
  - Have a lower mortality rate than older people in the absence of exposure: mortality rate is 4 times greater for old than for young

Age-Specific Death Rates

	Aluminum Reduction Plant Workers				Regional Sample of Non-Plant Workers				IRR_i
	Person-Years	Deaths	Rate	%	Person-Years	Deaths	Rate	%	
	Person-Years	Cases	Rate	Weight	Person-Years	Cases	Rate	Weight	
Young	10,000	50	0.005	0.91	100,000	50	0.0005	0.33	10
Old	1,000	4	0.004	0.09	200,000	400	0.002	0.67	2
Total	11,000	54	0.005		300,000	450	0.002		

- $$IRR_{crude} = \frac{54}{11,000} \Big/ \frac{450}{300,000} = 3.27$$

- Age is also a modifier
  - For young people, the mortality rate was 10 times greater in the exposed compared to the unexposed
  - For old people, the mortality rate was only 2 times greater in the exposed compared to the unexposed.

Utility of Standardization

- A comparison of crude rates can be misleading because the comparison can be biased due to confounding, and difficult to interpret in the presence of effect measure modification
- While **stratum-specific** rate ratios provide information on stratum-specific associations, a summary measure is often helpful to estimate the expected average association in populations with a particular distribution of the stratification factor(s).
- In the presence of effect measure modification, an estimate weighted by the distribution of the data (e.g. Mantel-Haenszel) is not appropriate.
- Standardization can be used to control confounding, and still yield a meaningful summary estimate when effect measure modification is present
- The **standardized** rate ratio (or difference) is a weighted average of the stratum-specific rate ratios (or differences), with the weights taken from a **standard** distribution to allow estimation of the expected average association in a population with a particular distribution of the stratification factor(s).
- Standardization can also be used for other measures, such as cumulative incidence or odds

BREAK

Week 7: Overview of Standardization

Video 2: Overview of Approaches to 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

Data-driven Weighted Average vs. Standardization

	Data-driven (e.g. IRR_{MH})	Standardization (e.g. SMR)
Confounding	Controls for confounding	Controls for confounding
Effect Measure Modification	Assumes no effect measure modification across strata	Appropriate even in the presence of effect modification
Weights	Based on information in strata	Based on a standard population
Statistical Efficiency	Weights are statistically efficient but they are not explicit	Weights may be statistically inefficient but they are explicit
Generalizability	Results are comparable to other studies in the absence of effect measure modification (or with similar distribution of modifiers).	Results are comparable to other studies with a similar distribution of the modifiers as the chosen standard population. Thus, the result depends on the choice of standard (e.g. SMR vs. SRR).
If no effect measure modification, expected value of data-driven and standardized measures are the same		

Choices for Standard Population

Distribution of the stratification factor(s) in the

- Unexposed (direct standardization)
- Exposed (indirect standardization)
- Total study population (sum of the study populations or groups)
- Artificial population (e.g. 1000 subjects per stratum)
- Population from which the study groups originate
  - e.g. The population of the state, province or country where the study is conducted
  - e.g. When comparing occupational groups in residents of a metropolitan area, total metropolitan area working population can serve as the standard
- External reference population (e.g. US census; World population)

Szklo & Nieto. Epidemiology. Jones & Bartlett Learning; 3 edition (October 24, 2012). p238-239

Unified and Traditional Approaches

- Unified and traditional approaches are algebraically identical

- Unified Approach
  - Weighted average of stratum-specific measures of association

- Traditional Approach
  - Standardize the rate in the exposed and standardize the rate in the unexposed
  - Calculate the ratio or difference of the standardized rate in the exposed and the standardized rate in the unexposed

BREAK

Week 7: Overview of Standardization

Video 3: Unified Approach:

Weighted Average of Stratum-Specific Associations

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

HARVARD T.H. CHAN
SCHOOL OF PUBLIC HEALTH

Unified Approach to Standardization

In the unified approach, the standardized rate ratio can be calculated as a weighted average of the stratum-specific rate ratios, similar to Mantel-Haenszel or inverse-variance summary estimates.

Direct vs. Indirect Standardization
Inference about Different Populations

- Weights are chosen to make inferences about the expected average associations in populations with a particular distribution of stratification factor(s)

- Typical choices of standards are often constructed to make inferences about a population with the distribution of the stratification factor(s) in
  - the **exposed** group in the study (“**indirect**” standardization)
  - the **unexposed** group in the study (“**direct**” standardization)
  - an external population of interest (“general population”, e.g. US census, world population)

Indirectly Standardized Rate Ratio (1)
Unified Approach

To make inferences about a population with the distribution of the stratification factors in the **exposed** group, the weights are the percent of the **exposed** person-time in stratum i multiplied by the rate in the unexposed:

Weights for indirectly standardized rate ratios: $w_{1i} = \left(\frac{N_{1i}}{N_1} \right) I_{0i}$

Indirectly Standardized Rate Ratio (2) Unified Approach

$$\frac{\sum w_i IRR_i}{\sum w_i} = \frac{\sum \left(\frac{N_{1i}}{N_1}\right)\left(I_{0i}\right)\left(IRR_i\right)}{\sum \left(\frac{N_{1i}}{N_1}\right)\left(I_{0i}\right)} = \frac{\sum \left(\frac{N_{1i}}{N_1}\right)\left(\cancel{I_{0i}}\right)\left(\frac{I_{1i}}{\cancel{I_{0i}}}\right)}{\sum \left(\frac{N_{1i}}{N_1}\right)\left(I_{0i}\right)} = \frac{\sum \left(\frac{N_{1i}}{N_1}\right) I_{1i}}{\sum \left(\frac{N_{1i}}{N_1}\right) I_{0i}}$$

Note that this can be conceptualized as a weighted average of stratum-specific rates

$$= \frac{\sum \left(\frac{\cancel{N_{1i}}}{N_1}\right)\left(\frac{a_i}{\cancel{N_{1i}}}\right)}{\sum \left(\frac{N_{1i}}{N_1}\right) I_{0i}} = \boxed{\frac{\sum \frac{a_i}{N_1}}{\sum \left(\frac{N_{1i}}{N_1}\right) I_{0i}}}$$

= observed rate in **exposed** / expected rate in **exposed** had they been **unexposed**

$$= \frac{\sum \frac{a_i}{N_1}}{\sum \left(\frac{N_{1i}}{N_1}\right) I_{0i}} = \frac{\cancel{\frac{1}{N_1}} \sum a_i}{\cancel{\frac{1}{N_1}} \left[\sum \left(N_{1i}\right) I_{0i}\right]} = \frac{\sum a_i}{\sum \left(N_{1i}\right) I_{0i}}$$

= observed number of **exposed** cases / expected number of **exposed** cases had they been **unexposed**

Directly Standardized Rate Ratio (1) Unified Approach

To make inferences about a population with the distribution of the stratification factors in the **unexposed** group, the weights are the percent of the **unexposed** person-time in stratum i multiplied by the rate in the unexposed:

Weights for directly standardized rate ratios: $w_{0i} = \left(\frac{N_{0i}}{N_0}\right) I_{0i}$

Directly Standardized Rate Ratio (2)
Unified Approach

$$\frac{\sum w_i IRR_i}{\sum w_i} = \frac{\sum \left(\frac{N_{0i}}{N_0}\right)\left(I_{0i}\right)\left(IRR_i\right)}{\sum \left(\frac{N_{0i}}{N_0}\right)\left(I_{0i}\right)} = \frac{\sum \left(\frac{N_{0i}}{N_0}\right)\left(\cancel{I_{0i}}\right)\left(\frac{I_{1i}}{\cancel{I_{0i}}}\right)}{\sum \left(\frac{N_{0i}}{N_0}\right)\left(I_{0i}\right)} = \frac{\sum \left(\frac{N_{0i}}{N_0}\right) I_{1i}}{\sum \left(\frac{N_{0i}}{N_0}\right) I_{0i}}$$

Note that this can be conceptualized as a weighted average of stratum-specific rates

$$= \frac{\sum \left(\frac{N_{0i}}{N_0}\right) I_{1i}}{\sum \left(\frac{\cancel{N_{0i}}}{N_0}\right)\left(\frac{b_i}{\cancel{N_{0i}}}\right)} = \boxed{\frac{\sum \left(\frac{N_{0i}}{N_0}\right) I_{1i}}{\sum \frac{b_i}{N_0}}}$$

$$= \frac{\text{expected rate in \textbf{unexposed} had they been \textbf{exposed}}}{\text{observed rate in \textbf{unexposed}}}$$

$$= \frac{\sum \left(\frac{N_{0i}}{N_0}\right) I_{1i}}{\sum \frac{b_i}{N_0}} = \frac{\frac{1}{\cancel{N_0}}\left[\sum \left(N_{0i}\right) I_{1i}\right]}{\frac{1}{\cancel{N_0}} \sum b_i} = \frac{\sum \left(N_{0i}\right) I_{1i}}{\sum b_i}$$

$$= \frac{\text{expected number of \textbf{unexposed} cases had they been \textbf{exposed}}}{\text{observed number of \textbf{unexposed} cases}}$$

Indirectly Standardized Rate Difference (1)
Unified Approach

To make inferences about a population with the distribution of the stratification factors in the **exposed** group, the weights are the percent of the **exposed** person-time in stratum i:

Weights for indirectly standardized rate differences: $w_{1i} = \dfrac{N_{1i}}{N_1}$

Indirectly Standardized Rate Difference (2) Unified Approach

$$\frac{\sum w_i IRD_i}{\sum w_i} = \frac{\sum w_i (I_{1i} - I_{0i})}{\sum w_i} = \frac{\sum w_i I_{1i} - w_i I_{0i}}{\sum w_i} = \frac{\sum w_i I_{1i}}{\sum w_i} - \frac{\sum w_i I_{0i}}{\sum w_i}$$

Note that this can be conceptualized as the difference between standardized rates

$$= \frac{\sum \frac{N_{1i}}{N_1} (I_{1i} - I_{0i})}{\sum \frac{N_{1i}}{N_1}} = \frac{\cancel{\frac{1}{N_1}} \sum N_{1i} (I_{1i} - I_{0i})}{\cancel{\frac{1}{N_1}} \sum N_{1i}} = \frac{\sum \left(\cancel{N_{1i}}\right)\left(\frac{a_i}{\cancel{N_{1i}}}\right)}{\sum N_{1i}} - \frac{\sum (N_{1i})(I_{0i})}{\sum N_{1i}}$$

$$= \boxed{\frac{\sum a_i}{\sum N_{1i}} - \frac{\sum (N_{1i})(I_{0i})}{\sum N_{1i}}}$$

= observed rate in **exposed** — expected rate in **exposed** had they been **unexposed**

Directly Standardized Rate Difference (1) Unified Approach

To make inferences about a population with the distribution of the stratification factors in the **unexposed** group, the weights are the percent of the **unexposed** person-time in stratum i:

Weights for directly standardized rate differences: $w_{0i} = \frac{N_{0i}}{N_0}$

Directly Standardized Rate Difference (2)
Unified Approach

$$\frac{\sum w_i IRD_i}{\sum w_i} = \frac{\sum w_i (I_{1i} - I_{0i})}{\sum w_i} = \frac{\sum w_i I_{1i} - w_i I_{0i}}{\sum w_i} = \frac{\sum w_i I_{1i}}{\sum w_i} - \frac{\sum w_i I_{0i}}{\sum w_i}$$

Note that this can be conceptualized as the difference between standardized rates

$$= \frac{\sum \frac{N_{0i}}{N_0}(I_{1i} - I_{0i})}{\sum \frac{N_{0i}}{N_0}} = \frac{\cancel{\frac{1}{N_0}} \sum N_{0i}(I_{1i} - I_{0i})}{\cancel{\frac{1}{N_0}} \sum N_{0i}} = \frac{\sum (N_{0i})(I_{1i})}{\sum N_{0i}} - \frac{\sum \left(\cancel{N_{0i}}\right)\left(\frac{b_i}{\cancel{N_{0i}}}\right)}{\sum N_{0i}}$$

$$= \boxed{\frac{\sum (N_{0i})(I_{1i})}{\sum N_{0i}} - \frac{\sum b_i}{\sum N_{0i}}}$$

= expected rate in **unexposed** had they been **exposed** — observed rate in **unexposed**

BREAK

Week 7: Overview of Standardization

Video 4: Traditional Approach:

Ratio or Difference of Standardized Rates

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

HARVARD T.H. CHAN
SCHOOL OF PUBLIC HEALTH

Traditional Approach to Standardization

- In the traditional approach, standardization can be accomplished by first standardizing the ***rates*** in the exposed and standardizing the ***rates*** in the unexposed groups
- The standardized rate ratio can be calculated by dividing these standardized rates
- The standardized rate difference can be calculated by subtracting these standardized rates
- This approach is mathematically identical to taking a weighted average of stratum-specific measures of association or using the inverse probability weighting approach

Direct vs. Indirect Standardization

Inference about Different Populations

	E+	E-
Cases	a_i	b_i
Person Time	N_{1i}	N_{0i}

$$\sum_{i=1}^{I} N_{1i} = N_1 \qquad \sum_{i=1}^{I} N_{0i} = N_0$$

Indirect Standardization

- Standardized Morbidity Ratio (SMR)
- Weights that reflect the **exposed**

$$w_{1i} = \frac{N_{1i}}{N_1} =$$ proportion of **exposed** person-time in stratum i

$$\sum_{i=1}^{I} w_{1i} = 1$$

Direct Standardization

- Standardized Rate Ratio (SRR)
- Weights that reflect the **unexposed**

$$w_{0i} = \frac{N_{0i}}{N_0} =$$ proportion of **unexposed** person-time in stratum i

$$\sum_{i=1}^{I} w_{0i} = 1$$

Indirectly Standardized Rate Ratio
Traditional Approach

$$\frac{\sum w_i I_{1i} \big/ \cancel{\sum w_i}}{\sum w_i I_{0i} \big/ \cancel{\sum w_i}} = \frac{\sum w_i I_{1i}}{\sum w_i I_{0i}} = \frac{\sum \left(\frac{N_{1i}}{N_1}\right) I_{1i}}{\sum \left(\frac{N_{1i}}{N_1}\right) I_{0i}}$$

$$= \frac{\sum \left(\frac{\cancel{N_{1i}}}{N_1}\right)\left(\frac{a_i}{\cancel{N_{1i}}}\right)}{\sum \left(\frac{N_{1i}}{N_1}\right) I_{0i}} = \boxed{\frac{\sum \frac{a_i}{N_1}}{\sum \left(\frac{N_{1i}}{N_1}\right) I_{0i}}} = \frac{\text{observed rate in } \textbf{exposed}}{\text{expected rate in } \textbf{exposed} \text{ had they been } \textbf{unexposed}}$$

$$= \frac{\sum \frac{a_i}{N_1}}{\sum \left(\frac{N_{1i}}{N_1}\right) I_{0i}} = \frac{\frac{1}{\cancel{N_1}} \sum a_i}{\frac{1}{\cancel{N_1}} \left[\sum (N_{1i}) I_{0i}\right]} = \frac{\sum a_i}{\sum (N_{1i}) I_{0i}} = \frac{\text{observed number of } \textbf{exposed} \text{ cases}}{\text{expected number of } \textbf{exposed} \text{ cases had they been } \textbf{unexposed}}$$

Directly Standardized Rate Ratio
Traditional Approach

$$\frac{\sum w_i I_{1i} \Big/ \cancel{\sum w_i}}{\sum w_i I_{0i} \Big/ \cancel{\sum w_i}} = \frac{\sum w_i I_{1i}}{\sum w_i I_{0i}} = \frac{\sum \left(\frac{N_{0i}}{N_0}\right) I_{1i}}{\sum \left(\frac{N_{0i}}{N_0}\right) I_{0i}}$$

$$= \frac{\sum \left(\frac{N_{0i}}{N_0}\right) I_{1i}}{\sum \left(\frac{\cancel{N_{0i}}}{N_0}\right)\left(\frac{b_i}{\cancel{N_{0i}}}\right)} = \boxed{\frac{\sum \left(\frac{N_{0i}}{N_0}\right) I_{1i}}{\sum \frac{b_i}{N_0}}}$$

= (expected rate in **unexposed** had they been **exposed**) / (observed rate in **unexposed**)

$$= \frac{\sum \left(\frac{N_{0i}}{N_0}\right) I_{1i}}{\sum \frac{b_i}{N_0}} = \frac{\cancel{\frac{1}{N_0}}\left[\sum (N_{0i}) I_{1i}\right]}{\cancel{\frac{1}{N_0}} \sum b_i} = \frac{\sum (N_{0i}) I_{1i}}{\sum b_i}$$

= (expected number of **unexposed** cases had they been **exposed**) / (observed number of **unexposed** cases)

Indirectly Standardized Rate Difference
Traditional Approach

$$=\frac{\sum w_{1i} I_{1i}}{\sum w_{1i}}-\frac{\sum w_{1i} I_{0i}}{\sum w_{1i}}=\frac{\sum\left(\frac{\cancel{N_{1i}}}{N_1}\right)\left(\frac{a_i}{\cancel{N_{1i}}}\right)}{\sum \frac{N_{1i}}{N_1}}-\frac{\sum\left(\frac{N_{1i}}{N_1}\right) I_{0i}}{\sum \frac{N_{1i}}{N_1}}$$

$$=\frac{\cancel{\frac{1}{N_1}}\sum a_i}{\cancel{\frac{1}{N_1}}\sum N_{1i}}-\frac{\cancel{\frac{1}{N_1}}\left[\sum\left(N_{1i}\right) I_{0i}\right]}{\cancel{\frac{1}{N_1}}\sum N_{1i}}$$

$$=\boxed{\frac{\sum a_i}{\sum N_{1i}}-\frac{\sum\left(N_{1i}\right)\left(I_{0i}\right)}{\sum N_{1i}}}$$

= observed rate in **exposed** — expected rate in **exposed** had they been **unexposed**

Directly Standardized Rate Difference
Traditional Approach

$$= \frac{\sum w_{0i} I_{1i}}{\sum w_{0i}} - \frac{\sum w_{0i} I_{0i}}{\sum w_{0i}} = \frac{\sum \left(\frac{N_{0i}}{N_0}\right) I_{1i}}{\sum \frac{N_{0i}}{N_0}} - \frac{\sum \left(\frac{\cancel{N_{0i}}}{N_0}\right)\left(\frac{b_i}{\cancel{N_{0i}}}\right)}{\sum \frac{N_{0i}}{N_0}}$$

$$= \frac{\cancel{\frac{1}{N_0}} \left[\sum (N_{0i}) I_{1i}\right]}{\cancel{\frac{1}{N_0}} \sum N_{0i}} - \frac{\cancel{\frac{1}{N_0}} \sum b_i}{\cancel{\frac{1}{N_0}} \sum N_{0i}}$$

$$= \boxed{\frac{\sum (N_{0i})(I_{1i})}{\sum N_{0i}} - \frac{\sum b_i}{\sum N_{0i}}}$$

= expected rate in **unexposed** had they been **exposed** — observed rate in **unexposed**

BREAK

Week 7: Overview of Standardization

Video 5: A Numerical Example

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

HARVARD T.H. CHAN
SCHOOL OF PUBLIC HEALTH

Aluminum Plant Exposure and Cancer Mortality

Indirectly Standardized Rate Ratio

	Aluminum Reduction Plant Workers				Regional Sample of Non-Plant Workers			
	N_{1i}	a_i	I_{1i}	N_{1i} / N_1	N_{0i}	b_i	I_{0i}	N_{0i} / N_0
	Person-Years	Cases	Rate	Weight	Person-Years	Cases	Rate	Weight
Young	10,000	50	0.005	0.91	100,000	50	0.0005	0.33
Old	1,000	4	0.004	0.09	200,000	400	0.002	0.67
Total	11,000	54			300,000	450		

- $$SMR = IRR_{indirect} = \frac{\sum \frac{a_i}{N_1}}{\sum \left(\frac{N_{1i}}{N_1}\right) I_{0i}} = \frac{\frac{54}{11{,}000}}{(0.91 \times 0.0005) + (0.09 \times 0.002)} = 7.71$$

- The **aluminum plant workers** have 7.71 times the rate of cancer deaths than they would have had if they had not been exposed to aluminum, assuming no residual confounding, no confounding by other factors, no selection bias and no information bias.

Aluminum Plant Exposure and Cancer Mortality
Directly Standardized Rate Ratio

	Aluminum Reduction Plant Workers				Regional Sample of Non-Plant Workers			
	N_{1i}	a_i	I_{1i}	N_{1i} / N_1	N_{0i}	b_i	I_{0i}	N_{0i} / N_0
	Person-Years	Cases	Rate	Weight	Person-Years	Cases	Rate	Weight
Young	10,000	50	0.005	0.91	100,000	50	0.0005	0.33
Old	1,000	4	0.004	0.09	200,000	400	0.002	0.67
Total	11,000	54			300,000	450		

- $$SRR = IRR_{direct} = \frac{\sum \left(\frac{N_{0i}}{N_0} \right) I_{1i}}{\sum \frac{b_i}{N_0}} = \frac{(0.33 \times 0.005) + (0.67 \times 0.004)}{\frac{450}{300,000}} = 2.89$$

- If the **non-plant workers** had received the aluminum plant exposure, the rate of cancer deaths would be 2.89 times the rate that would be seen if they had remained unexposed, assuming no residual confounding, no confounding by other factors, no selection bias and no information bias.

Aluminum Plant Exposure and Cancer Mortality

Indirectly Standardized Rate Difference

	Aluminum Reduction Plant Workers				Regional Sample of Non-Plant Workers			
	N_{1i}	a_i	I_{1i}	N_{1i} / N_1	N_{0i}	b_i	I_{0i}	N_{0i} / N_0
	Person-Years	Cases	Rate	Weight	Person-Years	Cases	Rate	Weight
Young	10,000	50	0.005	0.91	100,000	50	0.0005	0.33
Old	1,000	4	0.004	0.09	200,000	400	0.002	0.67
Total	11,000	54			300,000	450		

- $$\text{SMD} = \text{IRD}_{\text{indirect}} = \frac{\sum a_i}{\sum N_{1i}} - \frac{\sum (N_{1i})(I_{0i})}{\sum N_{1i}} = \frac{54}{11{,}000} - \frac{(10{,}000 \times 0.0005) + (1{,}000 \times 0.002)}{11{,}000}$$
  $$= 0.0049 - 0.0006 = 0.0043 \text{ cases / person} - \text{year}$$
- Among the **aluminum plant workers**, there was an excess of 4.3 cases of cancer deaths per 1,000 person-years compared to what they would have had if they had not been exposed to aluminum work, assuming no residual confounding, no confounding by other factors, no selection bias and no information bias.

Aluminum Plant Exposure and Cancer Mortality

Directly Standardized Rate Difference

	Aluminum Reduction Plant Workers				Regional Sample of Non-Plant Workers			
	N_{1i}	a_i	I_{1i}	N_{1i} / N_1	N_{0i}	b_i	I_{0i}	N_{0i} / N_0
	Person-Years	Cases	Rate	Weight	Person-Years	Cases	Rate	Weight
Young	10,000	50	0.005	0.91	100,000	50	0.0005	0.33
Old	1,000	4	0.004	0.09	200,000	400	0.002	0.67
Total	11,000	54			300,000	450		

- $$SRD = IRD_{direct} = \frac{\sum (N_{0i})(I_{1i})}{\sum N_{0i}} - \frac{\sum b_i}{\sum N_{0i}} = \frac{(100{,}000 \times 0.005) + (200{,}000 \times 0.004)}{300{,}000} - \frac{450}{300{,}000}$$
  $$= 0.0043 - 0.0015 = 0.0028 \text{ cases / person} - \text{year}$$

- If the **non-plant workers** were to receive the exposure(s) experienced by the aluminum workers, there would be an excess of 2.8 cases of cancer deaths per 1,000 person-years than would be seen if they had remained unexposed, assuming no residual confounding, no confounding by other factors, no selection bias and no information bias

BREAK

Week 7: Overview of Standardization

Video 6: Choosing a Standard

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

HARVARD T.H. CHAN
SCHOOL OF PUBLIC HEALTH

Aluminum Plant Exposure and Cancer Mortality
Age-Specific Incidence Rate Ratios

	Plant A				Plant B				Non-Plant Workers			
	Person-Years	Cases	Rate	Weight	Person-Years	Cases	Rate	Weight	Person-Years	Cases	Rate	Weight
Young	10,000	50	0.005	0.91	1,000	5	0.005	0.09	100,000	50	0.0005	0.33
Old	1,000	4	0.004	0.09	10,000	40	0.004	0.91	200,000	400	0.002	0.67
	11,000				11,000				300,000			

- Comparing Population A to the non-plant workers
  - In the young stratum, IRR = I_1/I_0 = 0.005/0.0005 = 10
  - In the old stratum, IRR = I_1/I_0 = 0.004/0.002 = 2
- Comparing Population B to the non-plant workers
  - In the young stratum, IRR = I_1/I_0 = 0.005/0.0005 = 10
  - In the old stratum, IRR = I_1/I_0 = 0.004/0.002 = 2

Rothman, K. J. 1986. Modern Epidemiology. Boston: Little, Brown. p.42-49

Aluminum Plant Exposure and Cancer Mortality

Indirect Standardization (SMR and SMD)

	Plant A				Plant B				Non-Plant Workers			
	Person-Years	Cases	Rate	Weight	Person-Years	Cases	Rate	Weight	Person-Years	Cases	Rate	Weight
Young	10,000	50	0.005	0.91	1,000	5	0.005	0.09	100,000	50	0.0005	0.33
Old	1,000	4	0.004	0.09	10,000	40	0.004	0.91	200,000	400	0.002	0.67
	11,000				11,000				300,000			

Compare Population A to Non-Plant Workers

SMR_A =7.71

SMD_A =0.0043 cases/person-year

Compare Population B to Non-Plant Workers

SMR_B =2.20

SMD_B =0.0022 cases/person-year

Aluminum Plant Exposure and Cancer Mortality
Direct Standardization (SRR and SRD)

	Plant A				Plant B				Non-Plant Workers			
	Person-Years	Cases	Rate	Weight	Person-Years	Cases	Rate	Weight	Person-Years	Cases	Rate	Weight
Young	10,000	50	0.005	0.91	1,000	5	0.005	0.09	100,000	50	0.0005	0.33
Old	1,000	4	0.004	0.09	10,000	40	0.004	0.91	200,000	400	0.002	0.67
	11,000				11,000				300,000			

Compare Plant A workers to Non-Plant Workers

SRR_A = 2.89

SRD_A =0.0028 cases/person-year

Compare Plant B workers to Non-Plant Workers

SRR_B = 2.89

SRD_B =0.0028 cases/person-year

Aluminum Plant Exposure and Cancer Mortality
Comparing Populations

- The SRRs used the common standard of the non-plant workers, so the rate ratios can validly be compared.
- When the age-specific incidence rates for the two cohorts are the same, the two SRRs will be the same no matter what the weights of the common standard.
- On the other hand, even if two populations have identical stratum-specific rates and therefore their directly standardized rates are identical, their indirectly standardized rates can be quite different.
- The two SMRs here were computed using different standards.
- In the presence of effect measure modification by the standardization factor(s), SMRs are not comparable unless the distribution of the stratification factor(s) among the exposed (i.e., the weights) is the same in the two populations.
- If there were no effect modification, however, any weighting scheme chosen, including the weighting scheme employed in indirect standardization, would yield estimates which would be comparable across groups.

Direct vs. Indirect Standardization
Advantages and Disadvantages

- **Direct (SRR)**
  - Requires subgroup-specific cumulative incidence/rate
  - Problematic when stratum-specific rate estimates are imprecise or unknown (e.g. company records only include total cases and person-time distribution)
  - Allows for comparison across study populations standardized to the same standard population

- **Indirect (SMR)**
  - Often used in occupational epidemiology
  - Particularly useful when
    - stratum-specific cumulative incidences/rates are missing in one of the groups under comparison
    - the study group(s) is (are) small so that the stratum-specific cumulative incidences/rates are unstable
  - Comparison of SMRs from different study populations is complicated by the fact that the weights used in obtaining SMRs are the stratum sizes of the exposed groups in the individual study populations rather than a common standard population

Indirect Standardization
Advantages

- Only need information on
  - Total number of exposed cases
  - Distribution of the stratification factor(s) among the exposed (to calculate weights)
  - Stratum-specific rates in the unexposed (in occupational health studies, often taken from the general population)

- The variance of the indirectly standardized estimate is much smaller than the variance of the directly standardized estimate
  - SMR: numerator is total number of exposed cases
  - SRR: stratum specific rates must be estimated

- Conceptually similar to the counterfactual contrast to infer the effect of the exposure among the exposed

BREAK

Week 7: Overview of Standardization

Video 7: Conditional Probabilities

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

HARVARD T.H. CHAN
SCHOOL OF PUBLIC HEALTH

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

- 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

A=exposure
Y=outcome
L=third variable

- **Pr[Y=1|A=a]=Σ_ℓPr[Y=1|L=ℓ,A=a]Pr[L=ℓ]**
  - where Σ_ℓ means sum over all values $\boldsymbol{\ell}$ 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 (2016). 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 (2016). Causal Inference.

BREAK

Week 7: Standardization: Inverse Probability Weighting (IPW)

Video 8: Motivation for Inverse Probability Weighting

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

HARVARD T.H. CHAN
SCHOOL OF PUBLIC HEALTH

Heart Transplant and Death
Hypothetical Randomized Experiment

- 2 million subjects with heart disease
  - numbers divided by 100,000 in example

- Variables:
  - A=1: heart transplant (exposure)
  - Y=1: death (outcome)
  - L=1: critical condition (prognostic factor)

- Goal: to estimate the effect of A on Y
  - calculate a cumulative incidence ratio

Exchangeability

- **Unconditional**
  - If the treated had been untreated, they would have been expected to have the same average outcome as the untreated
  - (and the other way around)

- **Conditional**
  - Same average outcome expected within each level of the covariates L

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

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

BREAK

Week 7: Standardization: Inverse Probability Weighting (IPW)

Video 9: How Inverse Probability Weighting Works

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

HARVARD T.H. CHAN
SCHOOL OF PUBLIC HEALTH

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

Hernán MA, Robins JM (2019). Causal Inference (forthcoming).

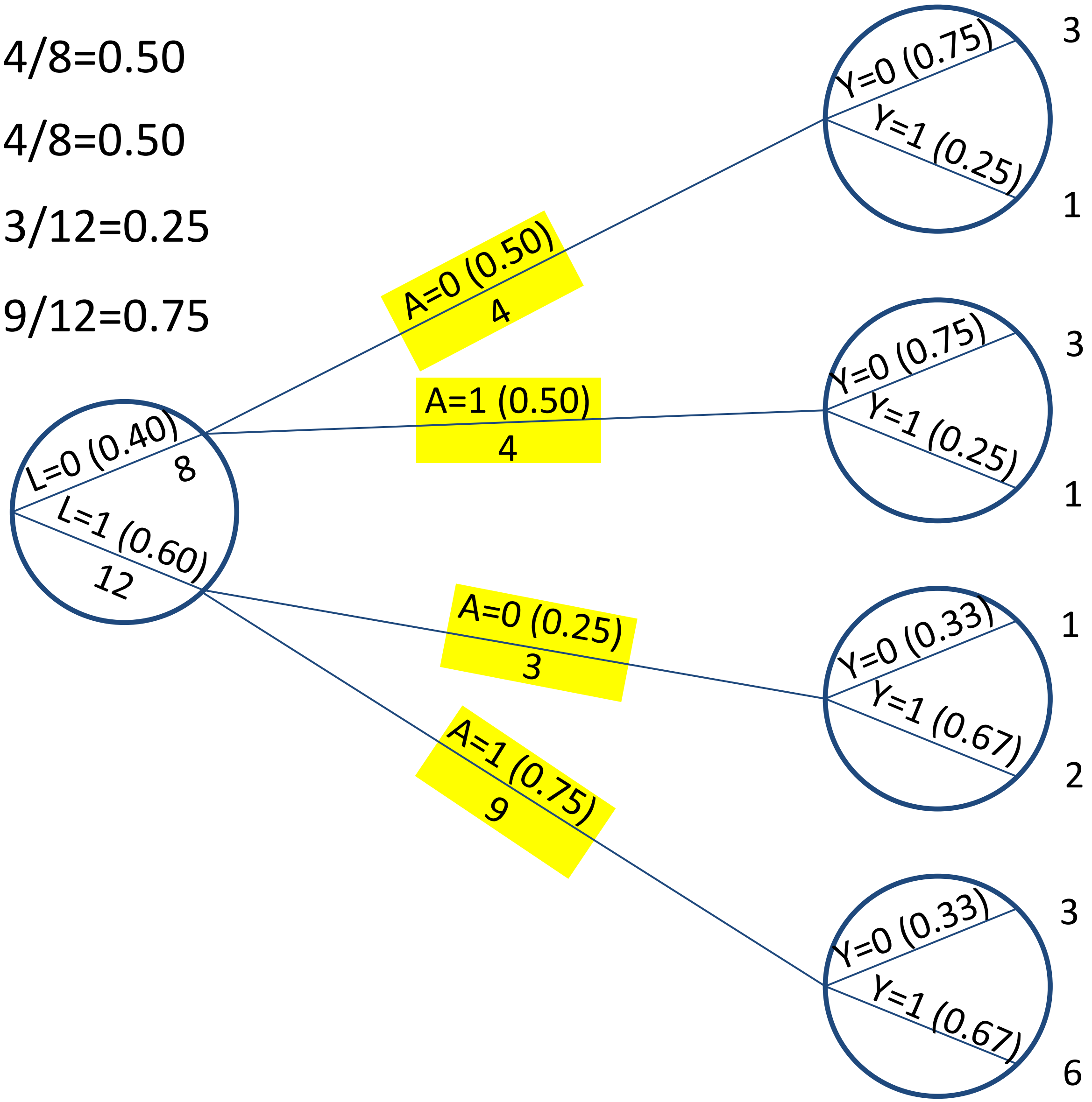
Heart Transplant and Death
Summarized in a Tree

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

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

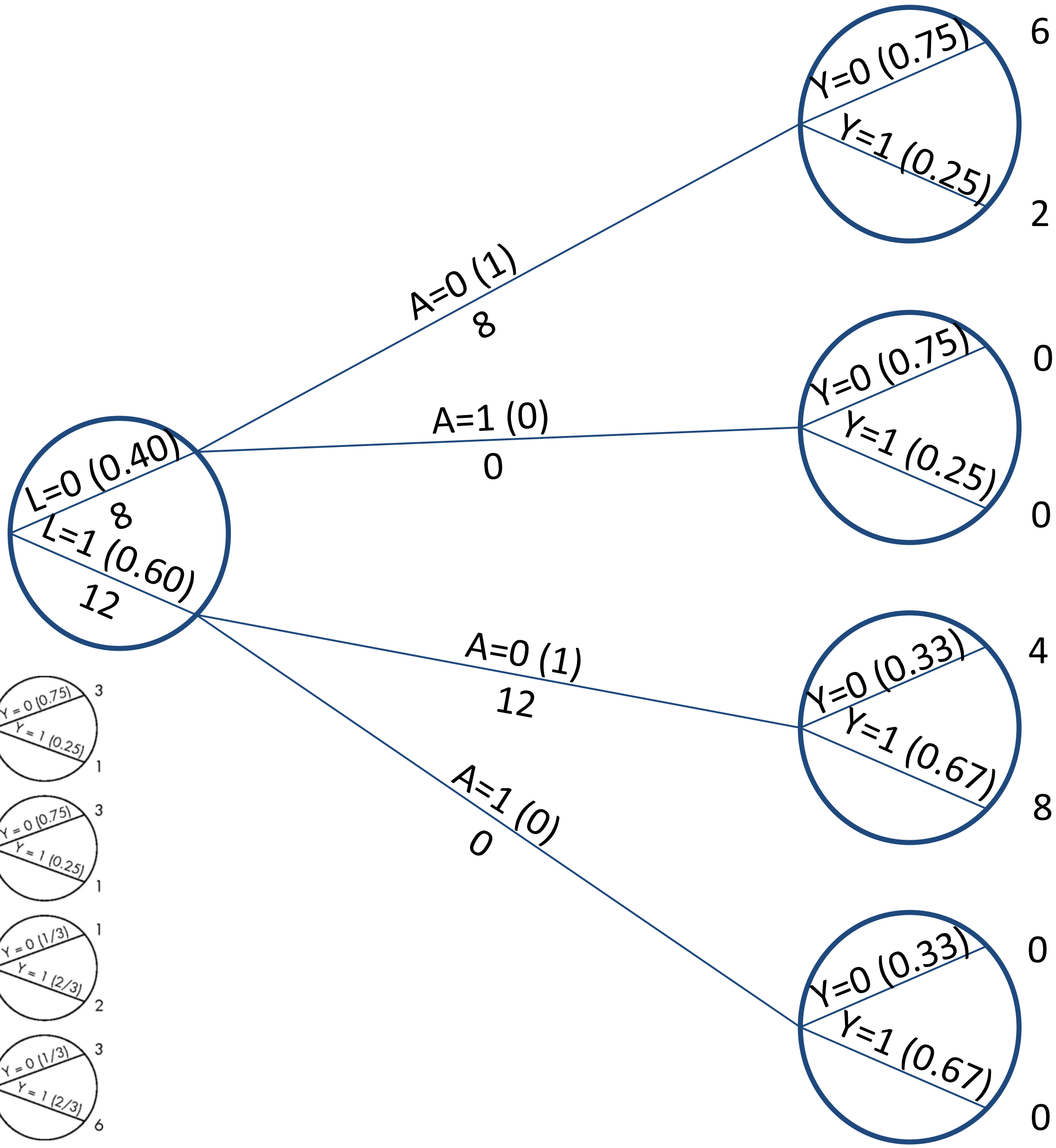
Pr[A=0|L=1] = 3/12=0.25

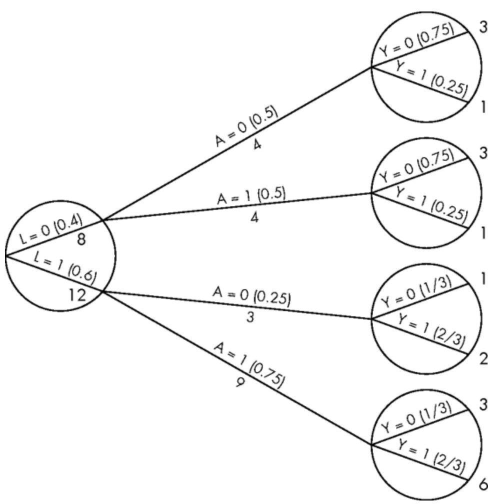
Pr[A=1|L=1] = 9/12=0.75



Heart Transplant and Death

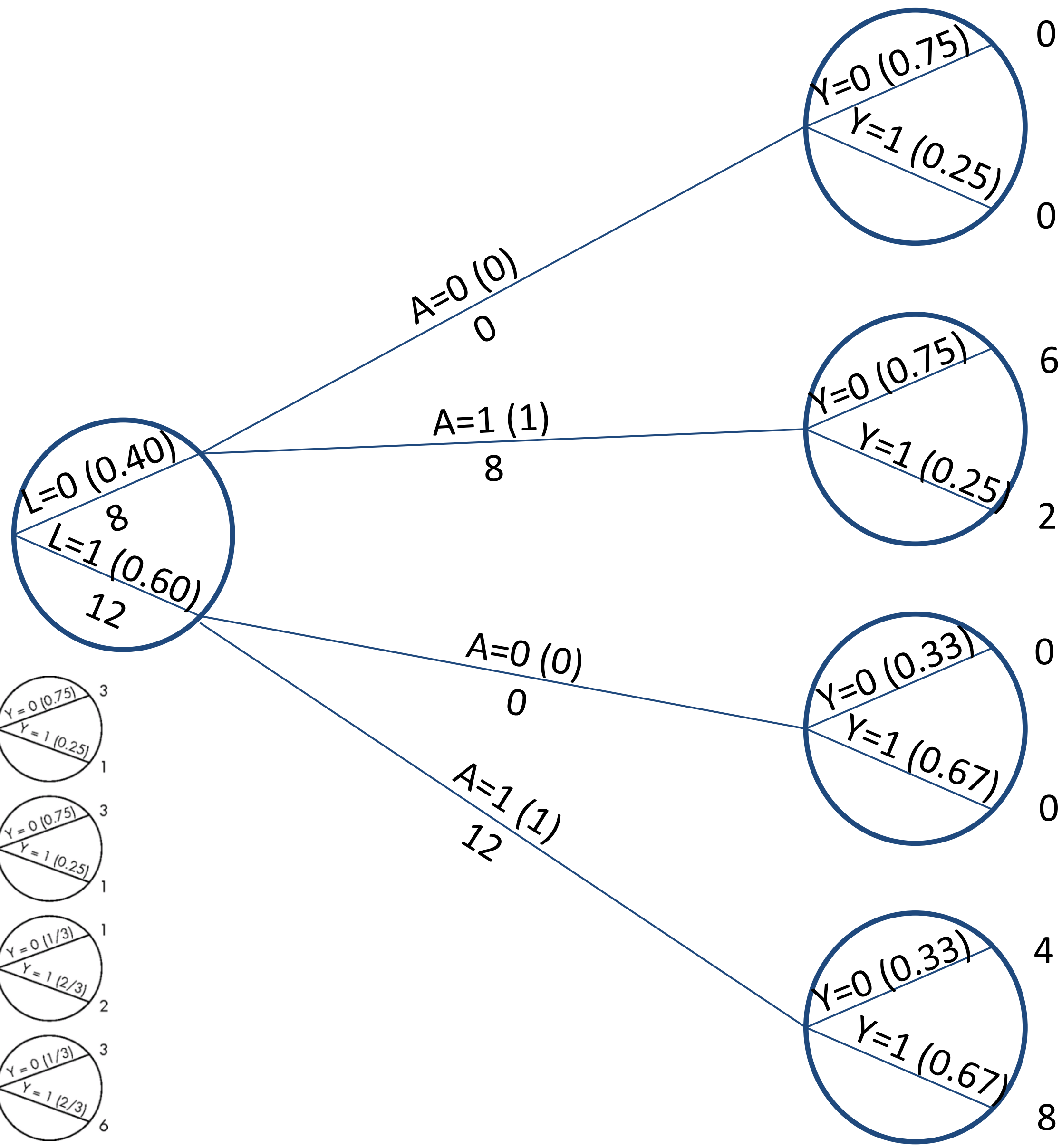
$\Pr[Y^{a=0}=1]$

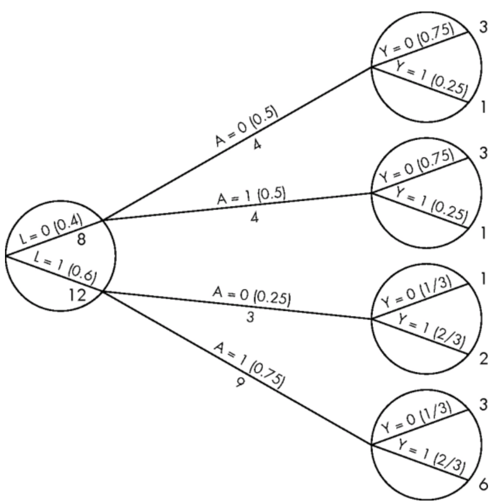




Heart Transplant and Death

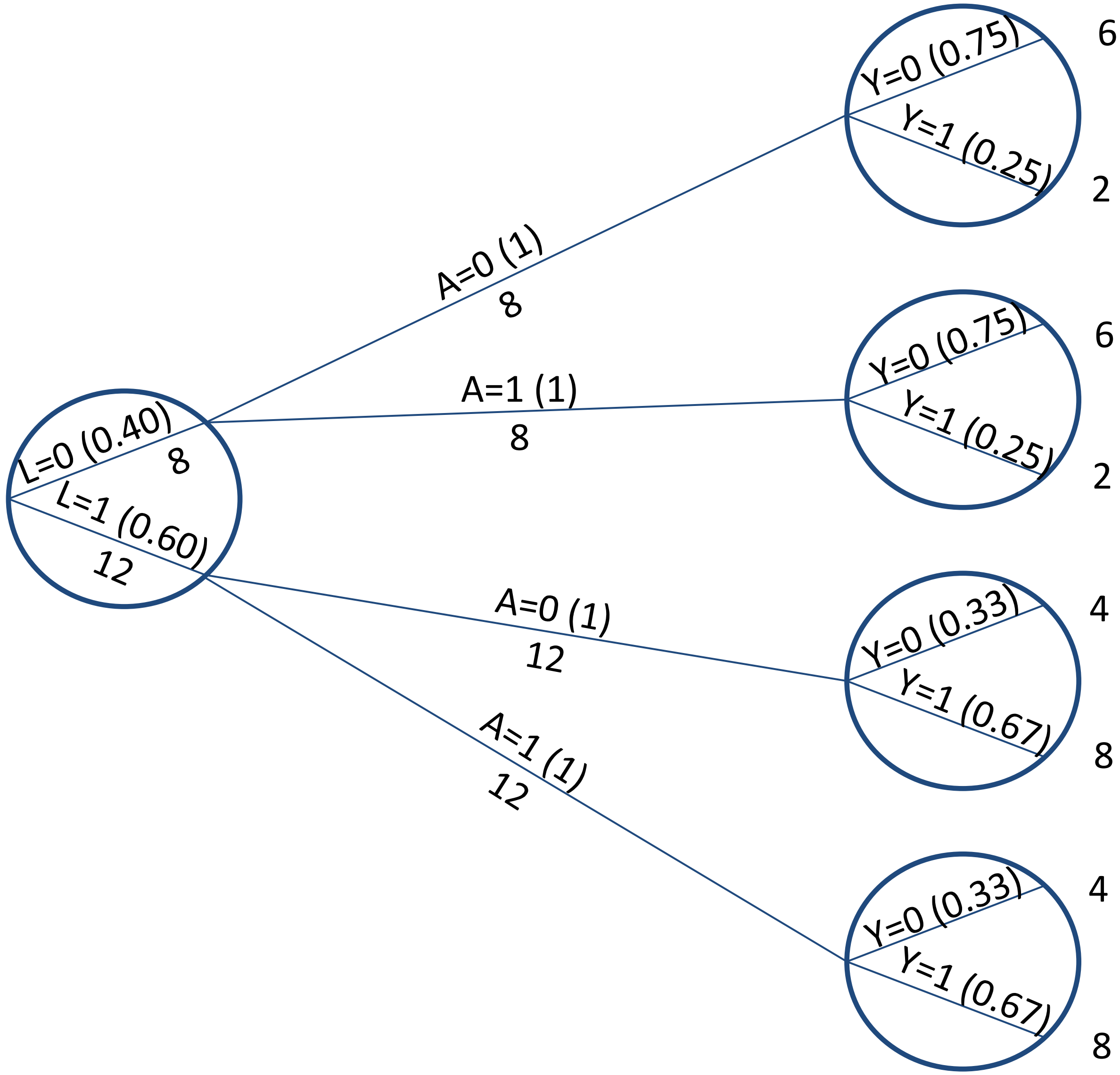
$Pr[Y^{a=1}=1]$



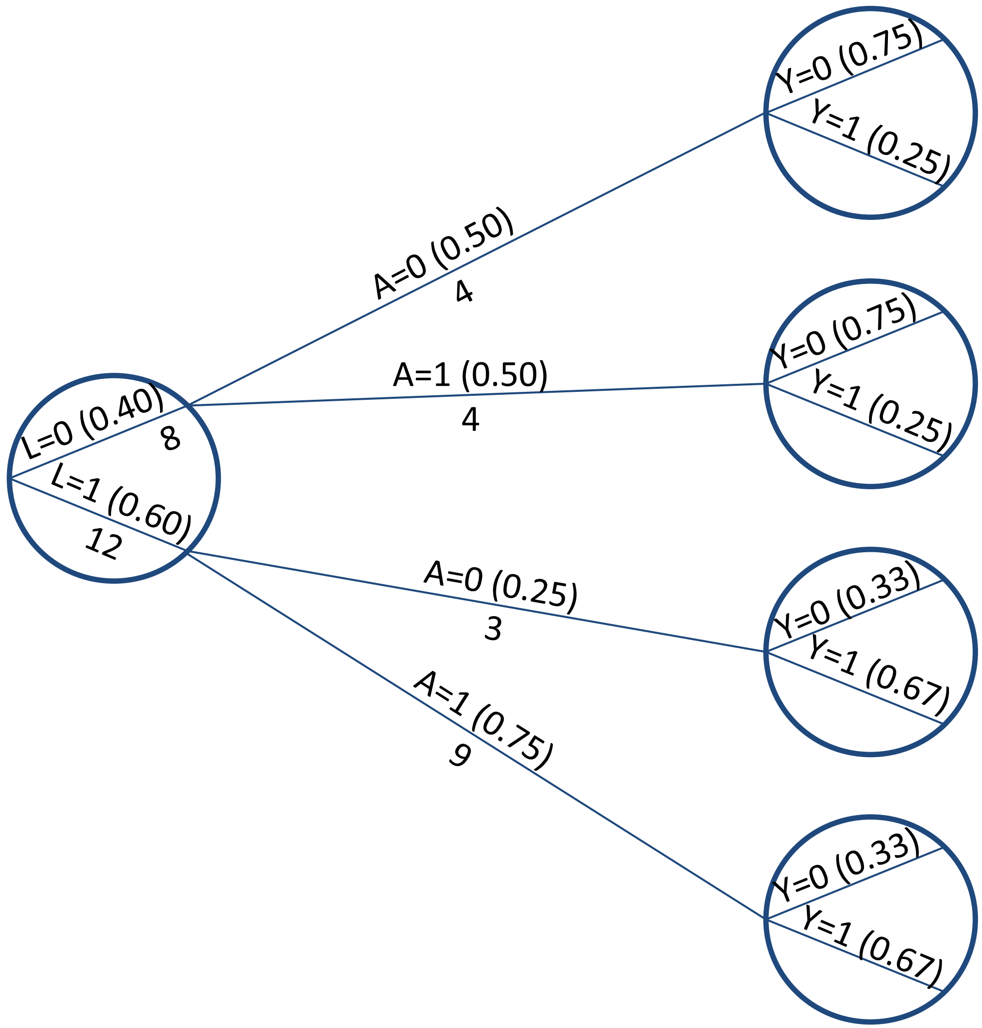


Heart Transplant and Death

Pseudo-Population for Pr[$Y^{a=0}$=1] and Pr[$Y^{a=1}$=1]



Heart Transplant and Death
Weights to Create Pseudo-Population for Total Population

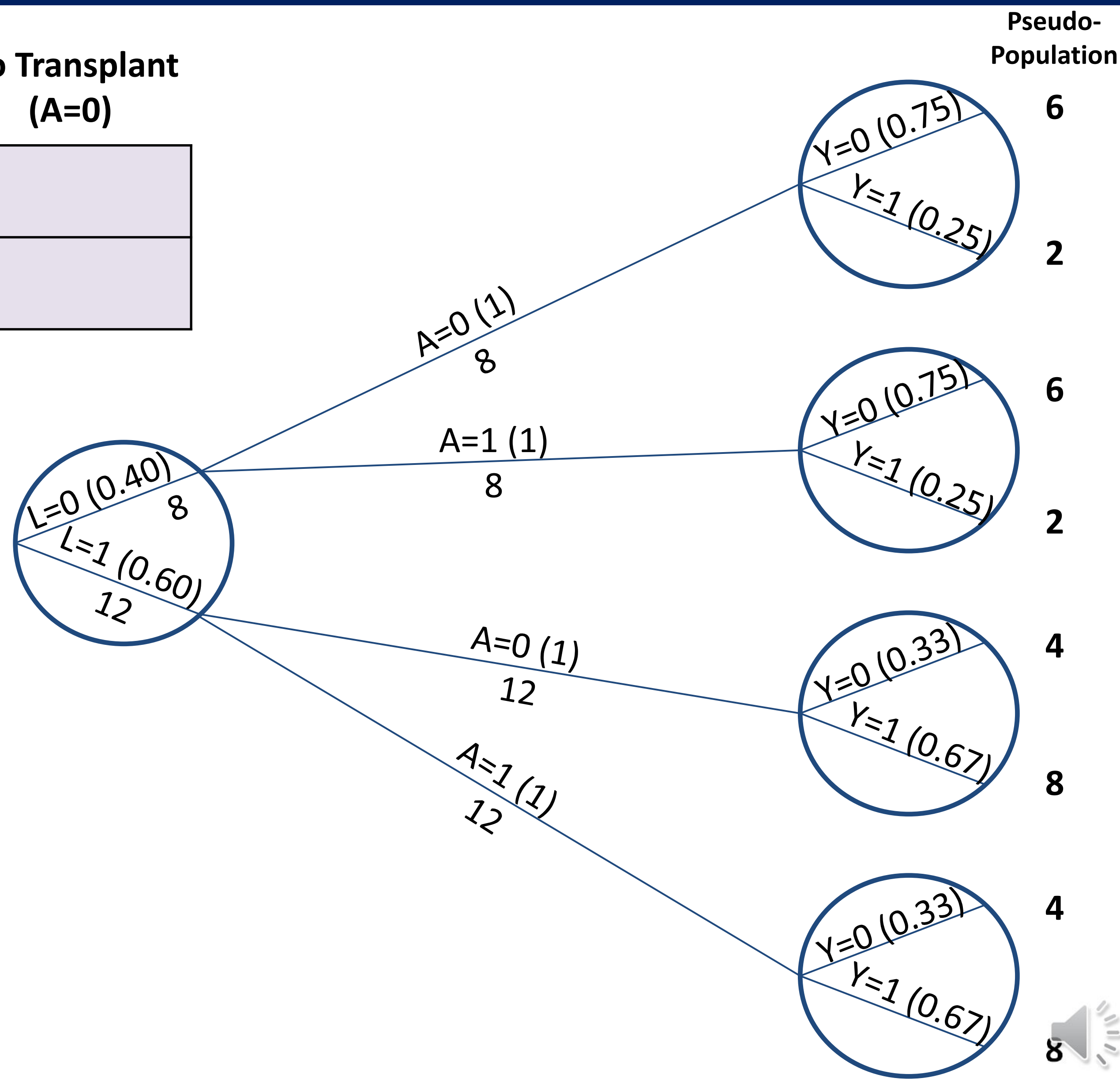


Observed Events	$\frac{1}{f(A\|L)}$	Pseudo-Population
3	1/0.5=2	6
1	1/0.5=2	2
3	1/0.5=2	6
1	1/0.5=2	2
1	1/0.25=4	4
2	1/0.25=4	8
3	1/0.75=1.33	4
6	1/0.75=1.33	8

Total **20** **40**

Heart Transplant and Death
Average Association for Total Population

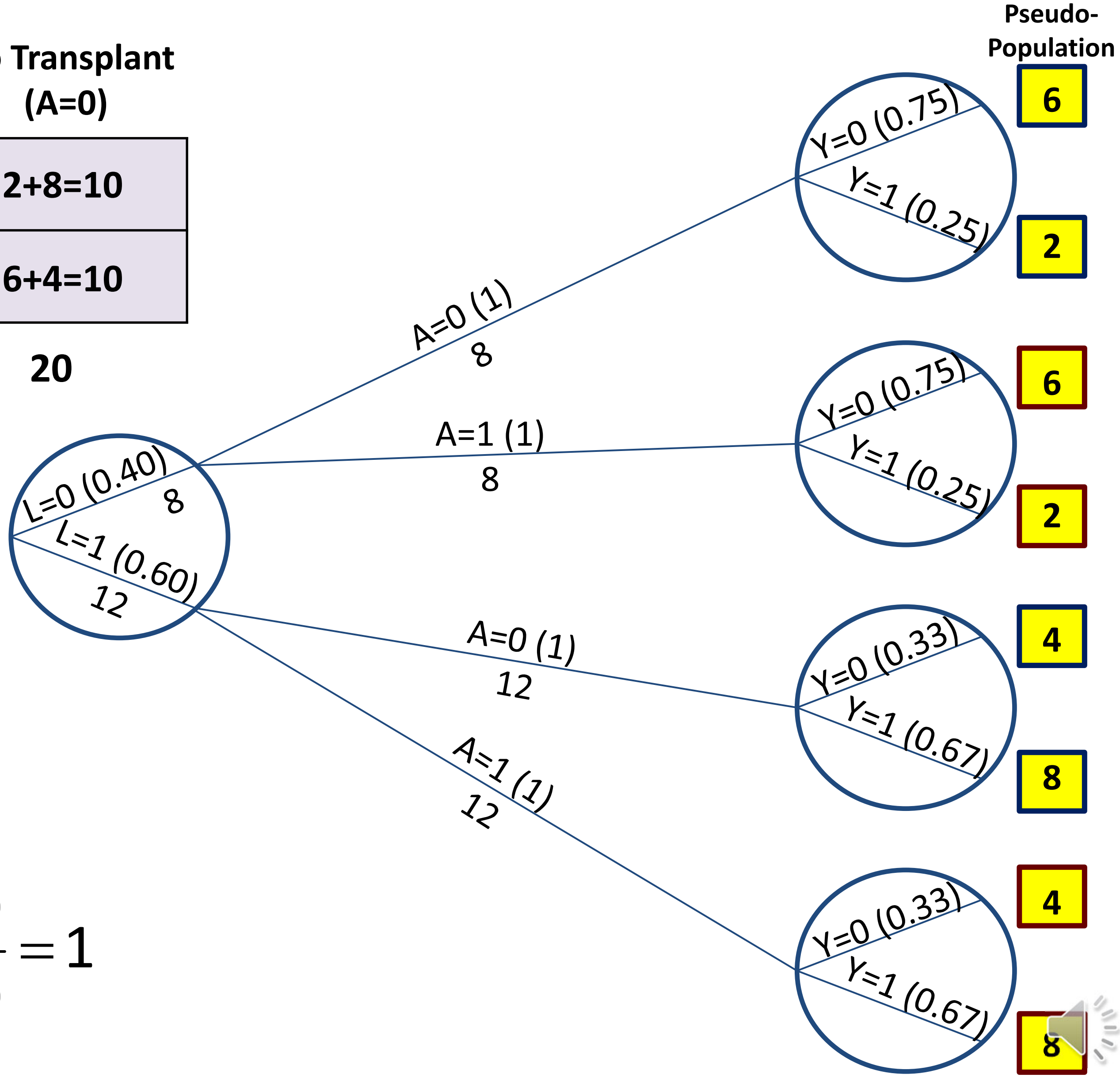
	Transplant (A=1)	No Transplant (A=0)
Death (Y=1)		
Survival (Y=0)		
Total		



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

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



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

BREAK

Week 7: Standardization: Inverse Probability Weighting (IPW)

Video 10: Inverse Probability Weighting Example

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

HARVARD T.H. CHAN
SCHOOL OF PUBLIC HEALTH

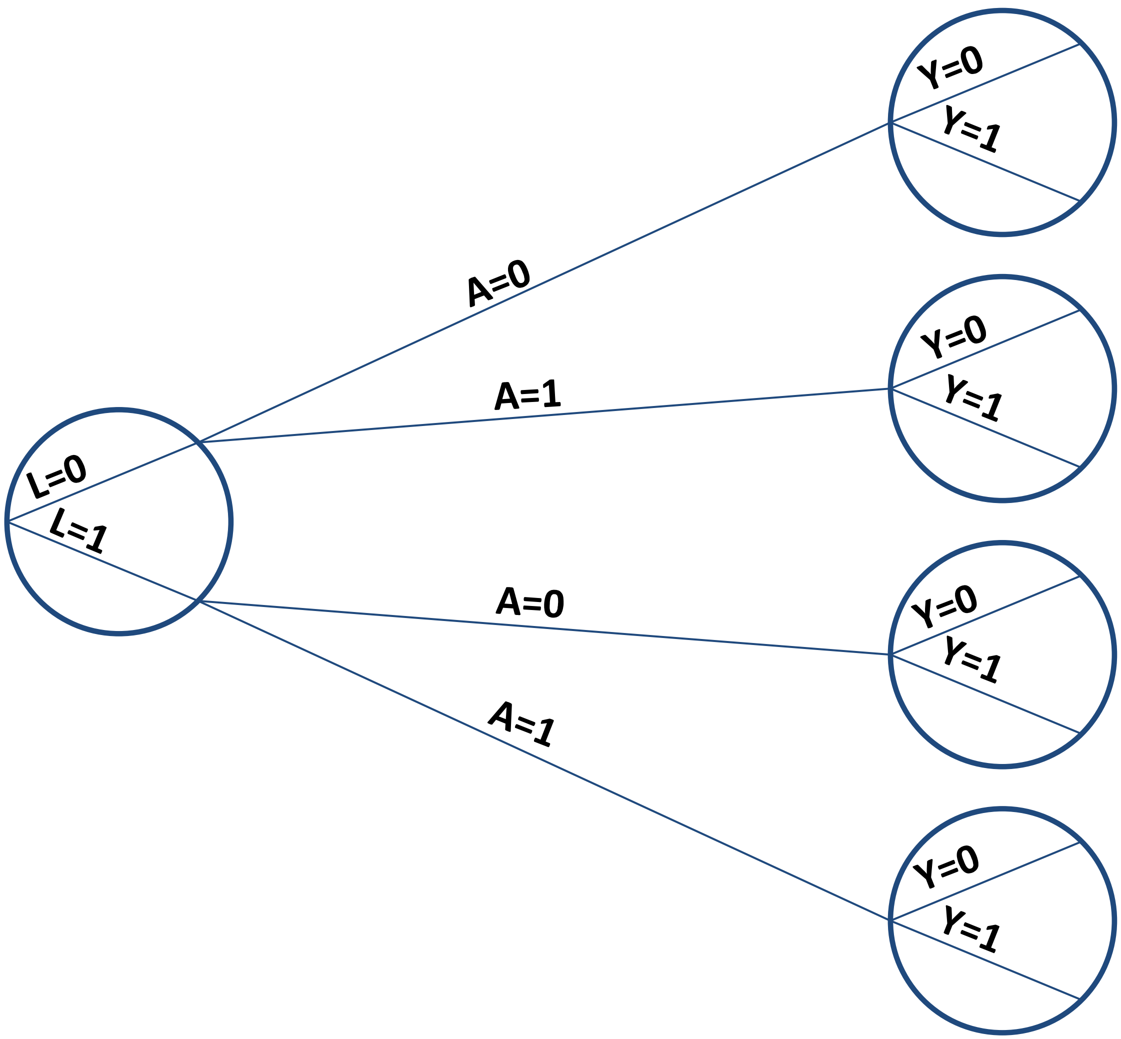
Jet Exhaust and Lung Cancer
Hypothetical Prospective Cohort Study

- Lung cancer is the leading cause of cancer death and the second most diagnosed cancer in both men and women in the United States.
- Airport workers are exposed to pollutants that increase cancer risk, e.g. nitrogen oxides, carbon dioxide, carbon monoxide, volatile organic compounds (VOCs) including polycyclic aromatic hydrocarbons (PAHs), sulfur dioxide, and fine and ultrafine particles (UFPs).
- Risk factors for lung cancer include
  - Smoking tobacco and being around others' smoke
  - Personal history (such as having radiation therapy)
  - Family history of lung cancer
- Goal: Compute the average association for the total population for the impact of jet exhaust exposure on the cumulative incidence of lung cancer
  - account for differences in the distribution of family history of lung cancer between the exposed and unexposed

Jet Exhaust and Lung Cancer
Summarized in a Table

	No Family History (L=0)			Family History (L=1)		
	Jet Exhaust (A=1)	No Jet Exhaust (A=0)	Total	Jet Exhaust (A=1)	No Jet Exhaust (A=0)	Total
Cancer (Y=1)	2	1	3	7	6	13
No Cancer (Y=0)	3	4	7	3	14	17
Total	5	5	10	10	20	30

Jet Exhaust and Lung Cancer
IPW Standardized to Total Population

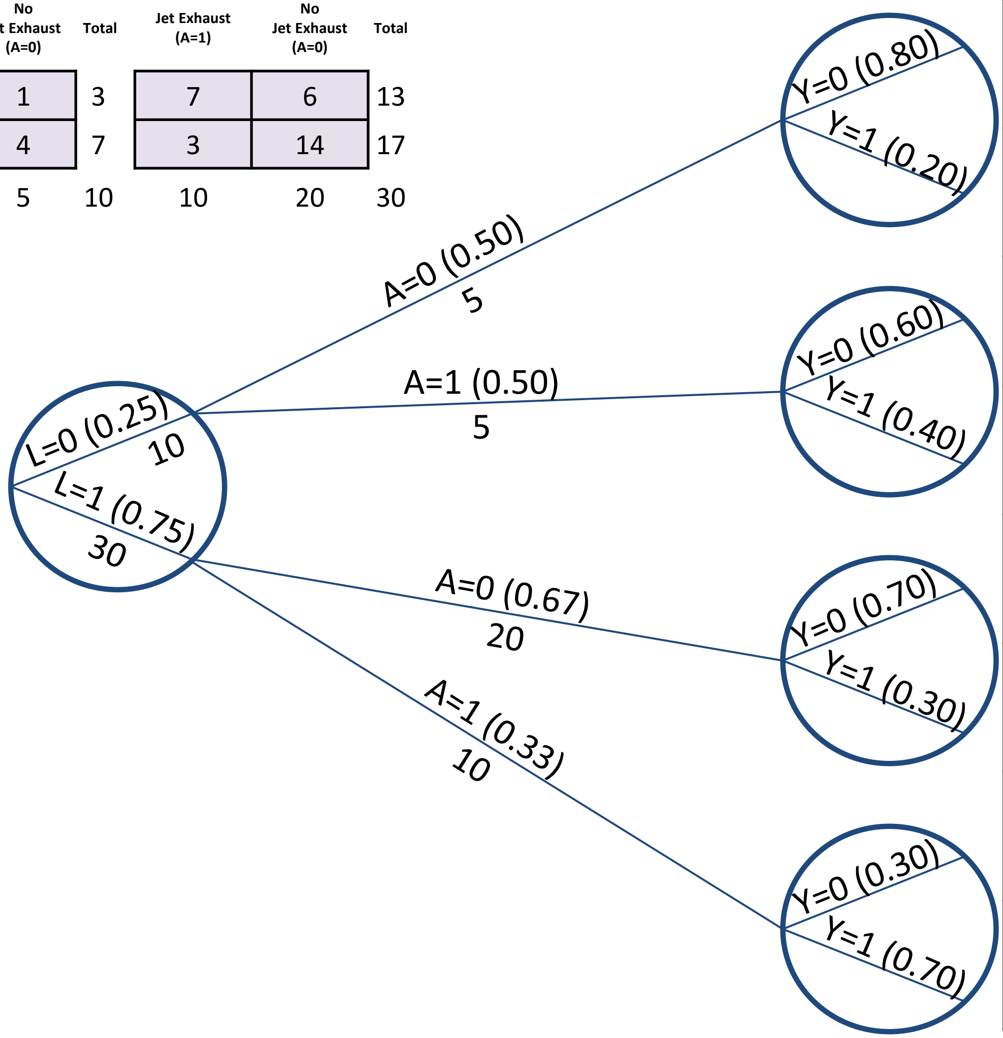


Observed Events	$\frac{1}{f(A\|L)}$	Pseudo-Population

Total

Jet Exhaust and Lung Cancer
IPW Standardized to Total Population

	No Family History (L=0)			Family History (L=1)		
	Jet Exhaust (A=1)	No Jet Exhaust (A=0)	Total	Jet Exhaust (A=1)	No Jet Exhaust (A=0)	Total
Cancer (Y=1)	2	1	3	7	6	13
No Cancer (Y=0)	3	4	7	3	14	17
Total	5	5	10	10	20	30



	Observed Events	$\frac{1}{f(A\|L)}$	Pseudo-Population
	4	1/0.5=2	8
	1	1/0.5=2	2
	3	1/0.5=2	6
	2	1/0.5=2	4
	14	1/0.67=1.5	21
	6	1/0.67=1.5	9
	3	1/0.33=3	9
	7	1/0.33=3	21
Total	**40**		**80**

Jet Exhaust and Lung Cancer
IPW Average Association for Total Population

	Pseudo-Population		
	Jet Exhaust (A=1)	**No Jet Exhaust (A=0)**	**Total**
Cancer (Y=1)			
No Cancer (Y=0)			
Total			

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

Jet Exhaust and Lung Cancer
IPW Average Association for Total Population

	Pseudo-Population		
	Jet Exhaust (A=1)	**No Jet Exhaust (A=0)**	**Total**
Cancer (Y=1)	4+21=**25**	2+9=**11**	**36**
No Cancer (Y=0)	6+9=**15**	8+21=**29**	**44**
Total	**40**	**40**	**80**

$$\frac{\Pr[Y^{a=1}=1]}{\Pr[Y^{a=0}=1]}=\frac{25/40}{11/40}=\frac{25}{11}=\frac{0.625}{0.275}=2.27$$

BREAK

Week 7: Standardization: Inverse Probability Weighting (IPW)

Video 11: Inverse Probability Weighting as Standardization

EPI202 – Epidemiologic Methods II

Murray A. Mittleman, MD, DrPH
Department of Epidemiology, Harvard TH Chan School of Public Hea

HARVARD T.H. CHAN
SCHOOL OF PUBLIC HEALTH

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]

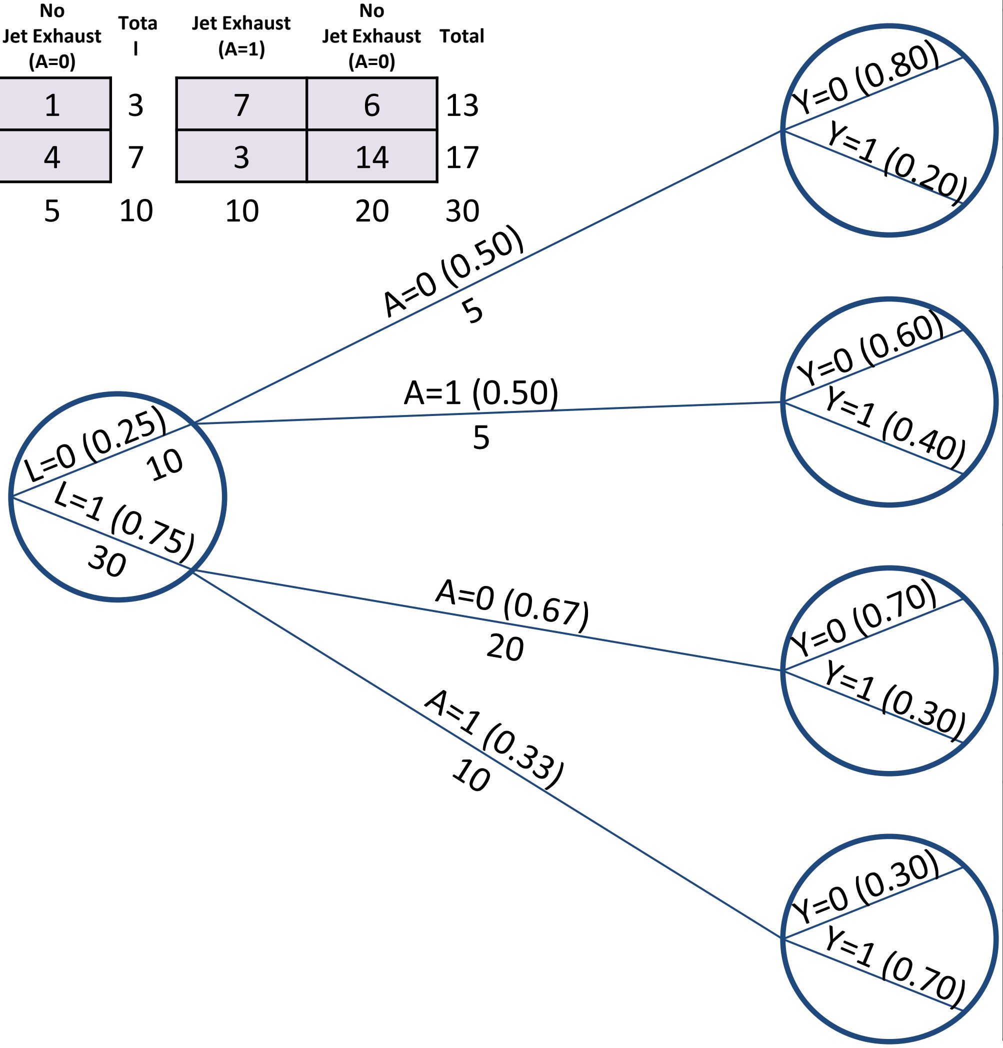
A=exposure
Y=outcome
L=stratification factor

- 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	$\frac{\Pr(Y^{a=1}=1 \mid A=0)}{\Pr(Y^{a=0}=1 \mid A=0)}$	$\frac{\Pr(A=0 \mid L)}{f(A \mid L)}$

Jet Exhaust and Lung Cancer
IPW Standardized to the Exposed

	No Family History (L=0)			Family History (L=1)		
	Jet Exhaust (A=1)	No Jet Exhaust (A=0)	Tota l	Jet Exhaust (A=1)	No Jet Exhaust (A=0)	Total
Cancer (Y=1)	2	1	3	7	6	13
No Cancer (Y=0)	3	4	7	3	14	17
Total	5	5	10	10	20	30



	Observed Events	Pr(A=1|L) / f(A|L)	Pseudo-Population
	4	0.5/0.5=1	4
	1	0.5/0.5=1	1
	3	0.5/0.5=1	3
	2	0.5/0.5=1	2
	14	0.33/0.67=0.5	7
	6	0.33/0.67=0.5	3
	3	0.33/0.33=1	3
	7	0.33/0.33=1	7
Total Exposed	**15**		**30**

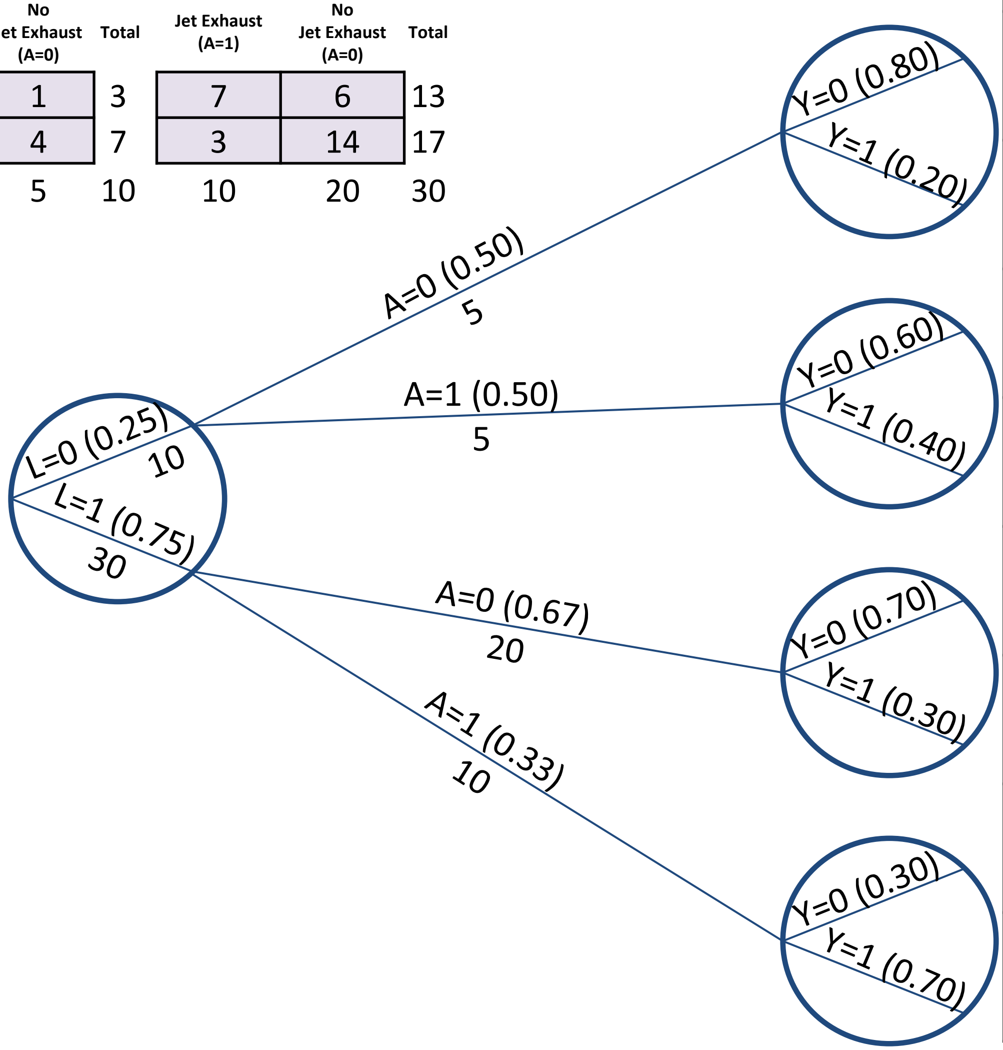
Jet Exhaust and Lung Cancer
IPW Average Association for the Exposed

	Pseudo-Population		
	Jet Exhaust (A=1)	No Jet Exhaust (A=0)	Total
Cancer (Y=1)	2+7=**9**	1+3=**4**	**13**
No Cancer (Y=0)	3+3=**6**	4+7=**11**	**17**
Total	**15**	**15**	**30**

$$\frac{\Pr[Y^{a=1}=1 \mid A=1]}{\Pr[Y^{a=0}=1 \mid A=1]} = \frac{9/15}{4/15} = \frac{9}{4} = \frac{0.60}{0.267} = 2.25$$

Jet Exhaust and Lung Cancer
IPW Standardized to the Unexposed

	No Family History (L=0)			Family History (L=1)		
	Jet Exhaust (A=1)	No Jet Exhaust (A=0)	Total	Jet Exhaust (A=1)	No Jet Exhaust (A=0)	Total
Cancer (Y=1)	2	1	3	7	6	13
No Cancer (Y=0)	3	4	7	3	14	17
Total	5	5	10	10	20	30



	Observed Events	Pr(A=0|L) / f(A|L)	Pseudo-Population
	4	0.5/0.5=1	4
	1	0.5/0.5=1	1
	3	0.5/0.5=1	3
	2	0.5/0.5=1	2
	14	0.67/0.67=1	14
	6	0.67/0.67=1	6
	3	0.67/0.33=2	6
	7	0.67/0.33=2	14
Total Unexposed	**25**		**50**

Jet Exhaust and Lung Cancer
IPW Average Association for the Unexposed

Pseudo-Population

	Jet Exhaust (A=1)	No Jet Exhaust (A=0)	Total
Cancer (Y=1)	2+14=**16**	1+6=**7**	**23**
No Cancer (Y=0)	3+6=**9**	4+14=**18**	**27**
Total	**25**	**25**	**50**

$$\frac{\Pr[Y^{a=1}=1\,|\,A=0]}{\Pr[Y^{a=0}=1\,|\,A=0]}=\frac{16/25}{7/25}=\frac{16}{7}=\frac{0.64}{0.28}=2.29$$

Jet Exhaust and Lung Cancer
Standardization Using Conditional Probabilities

	No Family History (L=0)			Family History (L=1)		
	Jet Exhaust (A=1)	No Jet Exhaust (A=0)	Total	Jet Exhaust (A=1)	No Jet Exhaust (A=0)	Total
Cancer (Y=1)	2	1	3	7	6	13
No Cancer (Y=0)	3	4	7	3	14	17
Total	5	5	10	10	20	30

Standardization to Total Population

$\Pr[Y^{a=1}=1]=\Pr[Y=1|A=1, L=0]$**$\Pr[L=0]$**$+\Pr[Y=1|A=1,L=1]$**$\Pr[L=1]$**$=(2/5)*$**$(10/40)$**$+(7/10)*$**$(30/40)$** $= 0.63$ $=$**2.27**

$\Pr[Y^{a=0}=1]=\Pr[Y=1|A=0, L=0]$**$\Pr[L=0]$**$+\Pr[Y=1|A=0,L=1]$**$\Pr[L=1]$**$=(1/5)*$**$(10/40)$**$+(6/20)*$**$(30/40)$** $= 0.28$

Standardization to the Exposed

$\Pr[Y^{a=1}=1|A=1]=\Pr[Y=1|A=1, L=0]$**$\Pr[L=0|A=1]$**$+\Pr[Y=1|A=1,L=1]$**$\Pr[L=1|A=1]$**$=(2/5)*$**$(5/15)$**$+(7/10)*$**$(10/15)$** $= 0.60$ $=$**2.25**

$\Pr[Y^{a=0}=1|A=1]=\Pr[Y=1|A=0, L=0]$**$\Pr[L=0|A=1]$**$+\Pr[Y=1|A=0,L=1]$**$\Pr[L=1|A=1]$**$=(1/5)*$**$(5/15)$**$+(6/20)*$**$(10/15)$** $= 0.27$

Standardization to the Unexposed

$\Pr[Y^{a=1}=1|A=0]=\Pr[Y=1|A=1, L=0]$**$\Pr[L=0|A=0]$**$+\Pr[Y=1|A=1,L=1]$**$\Pr[L=1|A=0]$**$=(2/5)*$**$(5/25)$**$+(7/10)*$**$(20/25)$** $= 0.64$ $=$**2.29**

$\Pr[Y^{a=0}=1|A=0]=\Pr[Y=1|A=0, L=0]$**$\Pr[L=0|A=0]$**$+\Pr[Y=1|A=0,L=1]$**$\Pr[L=1|A=0]$**$=(1/5)*$**$(5/25)$**$+(6/20)*$**$(20/25)$** $= 0.28$

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

BREAK