Agenda

- Causal Effects
- Measures of Association
- Risk and Rate Ratios/Differences
- Odds Ratio
- Proportional Measures
- Number Needed to Treat
- Regression Coefficients
- Problems

Effect

- Consequence, Result, Outcome
- Would not have happened otherwise
- Quantification of effect of an intervention requires some estimate of what would have happened without the intervention

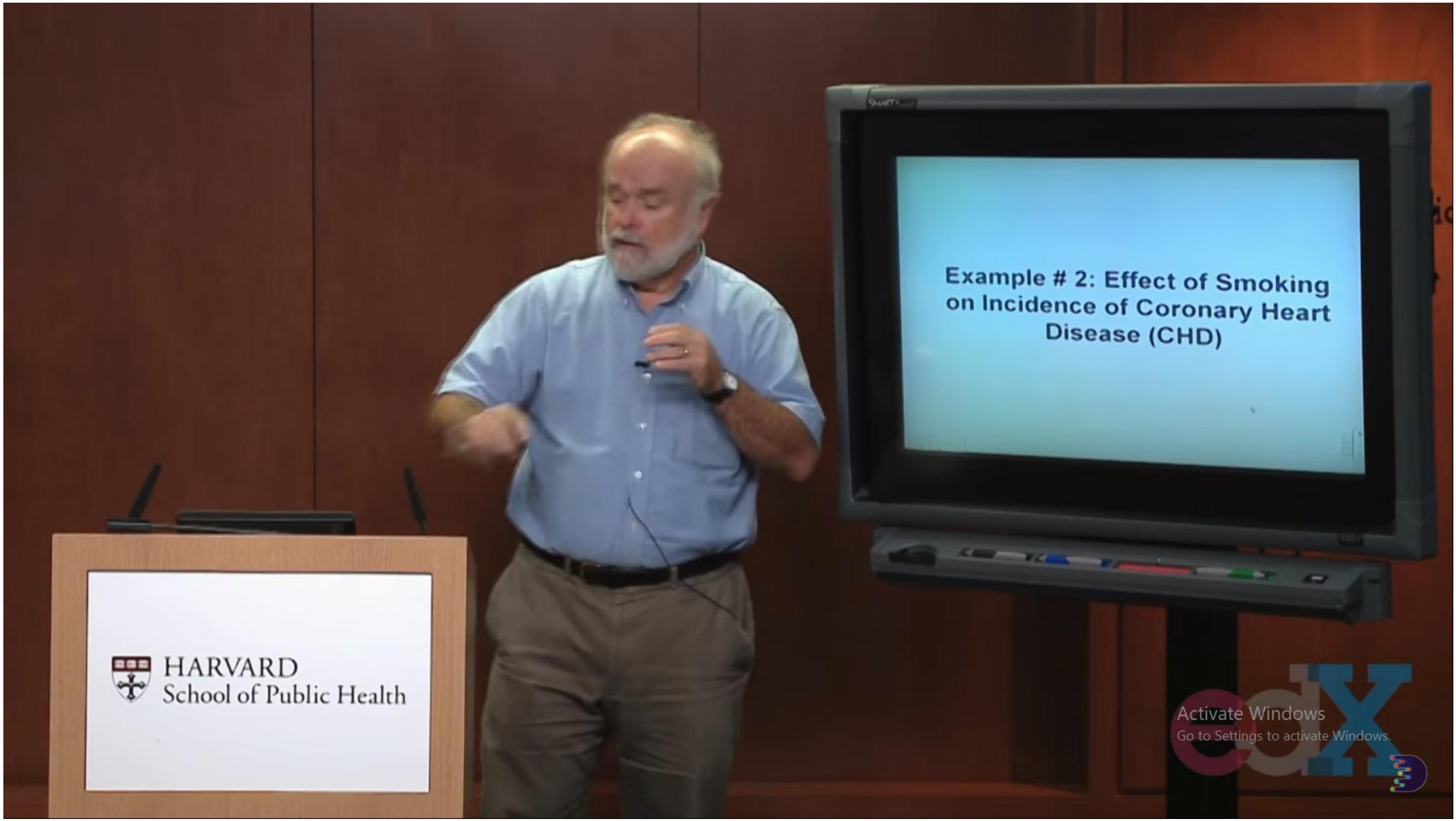
Need an estimate of what would have happened if Barack Obama was not elected as President

Option1

- ***"Are you better off than you were four years ago?"***
- Uses "pre-Obama-Presidency" (pre-intervention period) as point of reference
- **Problem**: Assumes there would have been change in outcomes over time

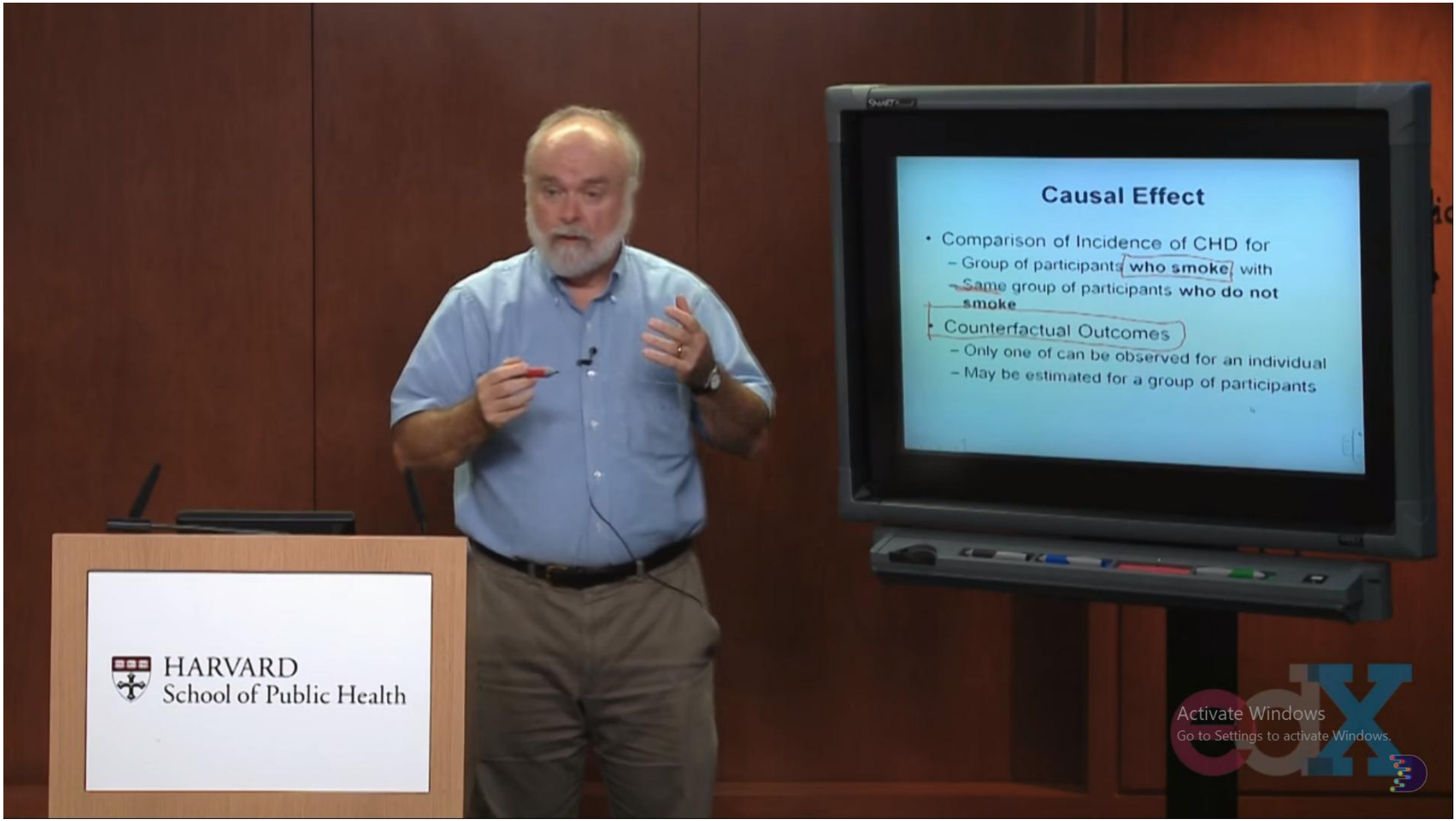
Option # 2

- *"What would things be like today if Barack Obama was not president?"*
- **Counterfactual Outcomes**
- **Problem**: Difficult to specify alternative (if Barack Obama was not president)
- **Problem**: Counterfactual Outcomes not observable


Example # 2: Effect of Smoking on Incidence of Coronary Heart Disease (CHD)
HARVARD
School of Public Health
Activate Windows
Go to Settings to activate Windows.

Causal Effect

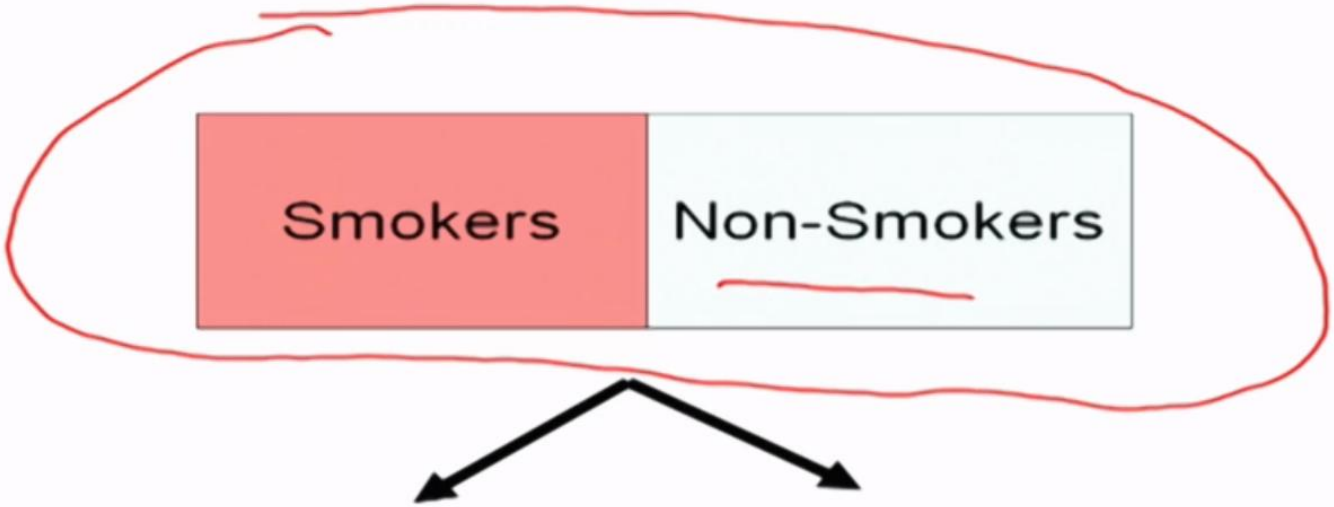
- Comparison of Incidence of CHD for
  - Group of participants **who smoke**, with
  - Same group of participants **who do not smoke**
- Counterfactual Outcomes
  - Only one of can be observed for an individual
  - May be estimated for a group of participants


Causal Effect
• Comparison of Incidence of CHD for
– Group of participants who smoke with
– Same group of participants who do not smoke
• Counterfactual Outcomes
– Only one of can be observed for an individual
– May be estimated for a group of participants
HARVARD
School of Public Health
Activate Windows
Go to Settings to activate Windows.

Estimated Causal Effect

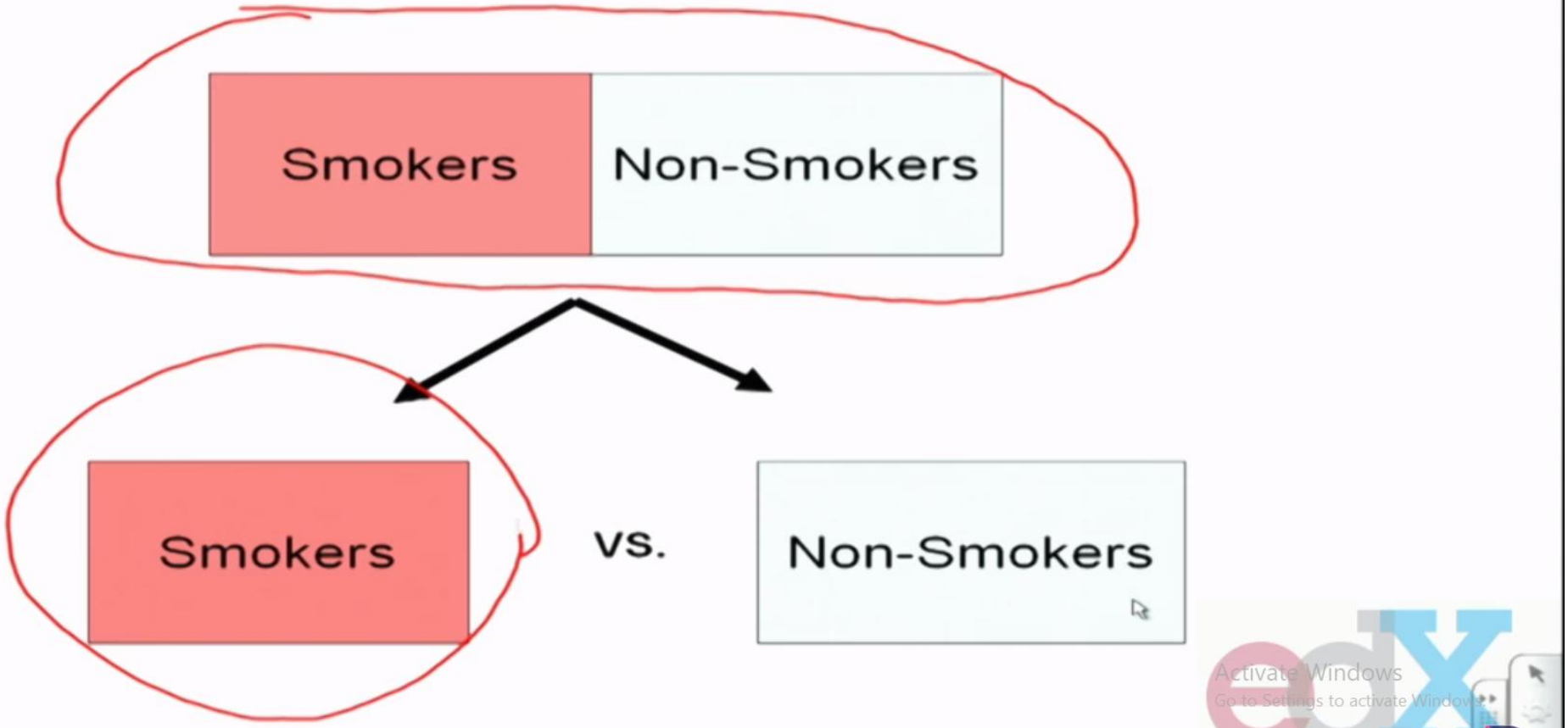
- Comparison of Incidence of CHD for
  - Group of participants **who smoke**, with
  - Different group of participants **who do not smoke**
- **Measure of Association**
- Estimates Causal Effect
  - Requires comparison groups be comparable (**free of confounding**)
  - Requires that study be **free of bias**

Effect of An Exposure



Based on slide from Miguel Hernan for EPI201 at HSP

Estimated Effect of An Exposure Measure of Association



Based on slide from Miguel Hernan for EPI201 at HSPH

Some Measures of Association

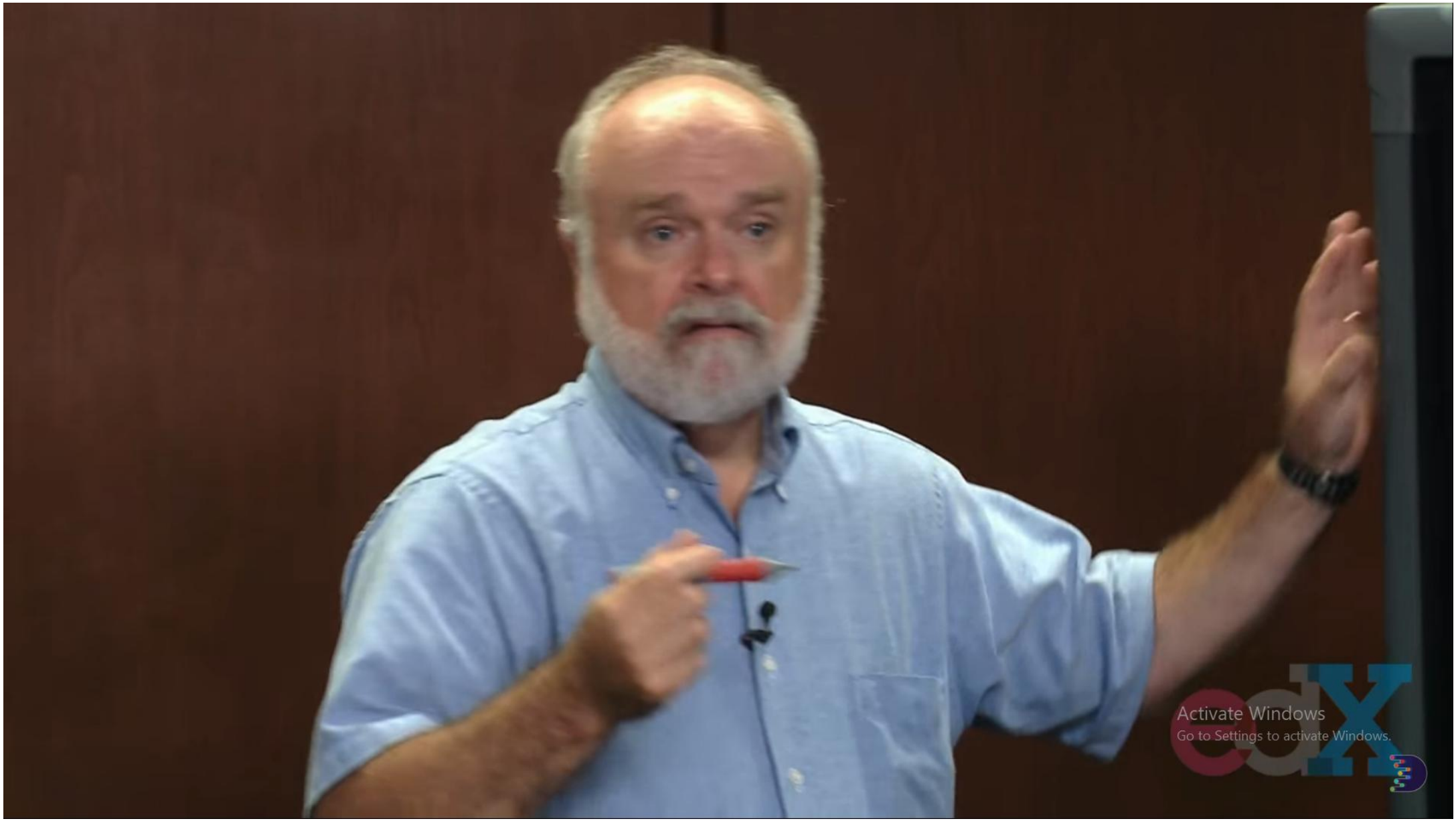
Outcome	Outcome Measure	Measure of Association
Binary	Proportion Odds Rates	**Risk Ratio/Difference** **Odds Ratio** **Rate Ratio/Difference**
Ordinal	Above Median/Percentile	**Above** **Ratio/Difference of Medians/Percentiles**
Continuous	Above Averages	**Above** **Ratio/Difference of Averages**

Measures of Association

- Compares outcome in exposed group (workshop attendees) to outcomes in non-exposed group (comparison group)
- Options:
  - **Ratio Measures**
  - **Difference Measures**

Binary Exposure and Binary Outcome

Exposure	Outcome +	Outcome -	Total	Estimated Risk
E+	A	B	N_1	A/N_1
E-	C	D	N_0	C/N_0


Activate Windows
Go to Settings to activate Windows.

Binary Exposure and Binary Outcome

Exposure	Outcome +	Outcome -	Total	Estimated Risk
E+	A	B	N_1	A/N_1
E-	C	D	N_0	C/N_0

Common Measures of Association

Measure of Association	Definition
Risk Ratio	$RR = R_1/R_0$
Risk Difference (Attributable Risk)	$RD = R_1 - R_0$
Disease Odds Ratio	$OR = \frac{R_1/1-R_1}{R_0/1-R_0}$

Example: Incidence of First Headaches (D) among 100 students over 7 days

ID#							Days of Follow-up
1.	D						1
2.	D						1
3.	D						1
4.	D						1
5.	X	D					2
6.	X	X	D				3
7.	X	X	D				3
8.	X	X	D				3
9.	X	X	D				3
10.	X	X	X	D			4
11.	X	X	X	D			4
12.	X	X	X	X	D		5
13.	X	X	X	X	X	D	6
14.	X	X	X	X	X	D	6
15.	X	X	X	X	X	X	D 7
16.	X	X	X	X	X	X	X 7
...							
100.	X	X	X	X	X	X	X 7
Day	1	2	3	4	5	6	7

Example: Incidence of First Headaches (D) among 50 Men

ID#								Days of Follow-up
1.	D							1
2.	D							1
3.	D							1
4.	D							1
5.	X	D						2
6.	X	X	D					3
7.	X	X	D					3
8.	X	X	D					3
9.	X	X	D					3
10.	X	X	X	D				4
11.	X	X	X	D				4
12.	X	X	X	X	D			5
13.	X	X	X	X	X	D		6
14.	X	X	X	X	X		D	6
15.	X	X	X	X	X	X	D	7
16.	X	X	X	X	X	X	X	7
...								
100.	X	X	X	X	X	X	X	7
Day	1	2	3	4	5	6	7	

Example: Incidence of First Headaches (D) among 50 Women

ID#								Days of Follow-up
1.	D							1
2.	D							1
3.	D							1
4.	D							1
5.	X	D						2
6.	X	X	D					3
7.	X	X	D					3
8.	X	X	D					3
9.	X	X	D					3
10.	X	X	X	D				4
11.	**X**	**X**	**X**	**D**				**4**
12.	**X**	**X**	**X**	**X**	**D**			**5**
13.	**X**	**X**	**X**	**X**	**X**	**D**		**6**
14.	**X**	**X**	**X**	**X**	**X**	**D**		**6**
15.	**X**	**X**	**X**	**X**	**X**	**X**	**D**	**7**
16.	X	X	X	X	X	X	X	7
...								
100.	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**7**
Day	1	2	3	4	5	6	7	Σ

Cumulative Incidence Data

	Headache	No Headache	Total	Estimated Risk
Males	10	40	50	10/50 =.20
Females	5	45	50	5/50 = .10

Risk Ratio = 10/50 / 5/50 = .20 / .10 = 2.0

Risk Difference = .20 - .10 = .10

Cumulative Incidence Data

	Headache	No Headache	Total	Estimated Risk
Males	10	40	50	10/50 =.20
Females	5	45	50	5/50 = .10

Risk Ratio = 10/50 / 5/50 = .20 / .10 = 2.0

Risk Difference = .20 - .10 = .10

Cumulative Incidence Data

	Headache	No Headache	Total	Estimated Risk
Males	10	40	50	10/50 =.20
Females	5	45	50	5/50 = .10

Risk Ratio = 10/50 / 5/50 = .20 / .10 = 2.0

Risk Difference = .20 - .10 = .10

Incidence Rate Data

	Headache	Person-Days of Follow-up	Incidence Rate
Males	10	302	10/302pd
Females	5	343	5/343pd

Rate Ratio = (10/302pd) / (5/343pd) = 2.27

Rate Difference = (10/302pd) - (5/343pd)

= 3.31/100pd – 1.46/100pd = 1.85/100pd

Incidence Rate Data

	Headache	Person-Days of Follow-up	Incidence Rate
Males	10	302	10/302pd
Females	5	343	5/343pd

Rate Ratio = (10/302pd) / (5/343pd) = 2.27

Rate Difference = (10/302pd) - (5/343pd)

= 3.31/100pd – 1.46/100pd = 1.85/100pd

Incidence Rate Data

	Headache	Person-Days of Follow-up	Incidence Rate
Males	10	302	10/302pd
Females	5	343	5/343pd

Rate Ratio = (10/302pd) / (5/343pd) = 2.27

Rate Difference = (10/302pd) - (5/343pd)

= 3.31/100pd – 1.46/100pd = 1.85/100pd

Cumulative Incidence Data

	Headache	No Headache	Total	**Disease Odds**
Males	10	40	50	
Females	5	45	50	

Cumulative Incidence Data

	Headache	No Headache	Total	Disease Odds
Males	10	40	50	10/40
Females	5	45	50	

Cumulative Incidence Data

	Headache	No Headache	Total	Disease Odds
Males	10	40	50	10/40 = .25
Females	5	45	50	5/45 = .11

Cumulative Incidence Data

	Headache	No Headache	Total	Disease Odds
Males	10	40	50	**10/40 = .25**
Females	5	45	50	**5/45 = .11**

Odds Ratio = (10/40) / (5/45) =2.25

Odds Ratio

- Effect measure of choice for **Case Control Studies**
  - Option for Cohort Studies with Cumulative Incidence data
- Effect measure obtained from **logistic regression models**
- Symmetrical

Odds Ratio

- Effect measure of choice for **Case Control Studies**
  - Option for Cohort Studies with Cumulative Incidence data
- Effect measure obtained from **logistic regression models**
- Symmetrical

Odds Ratio

- **Odds** for not developing headache = 1/(odds for developing disease)

Cumulative Incidence Data

	Headache	No Headache	Total	Non-Disease Odds
Males	10	40	50	40/10
Females	5	45	50	45/5

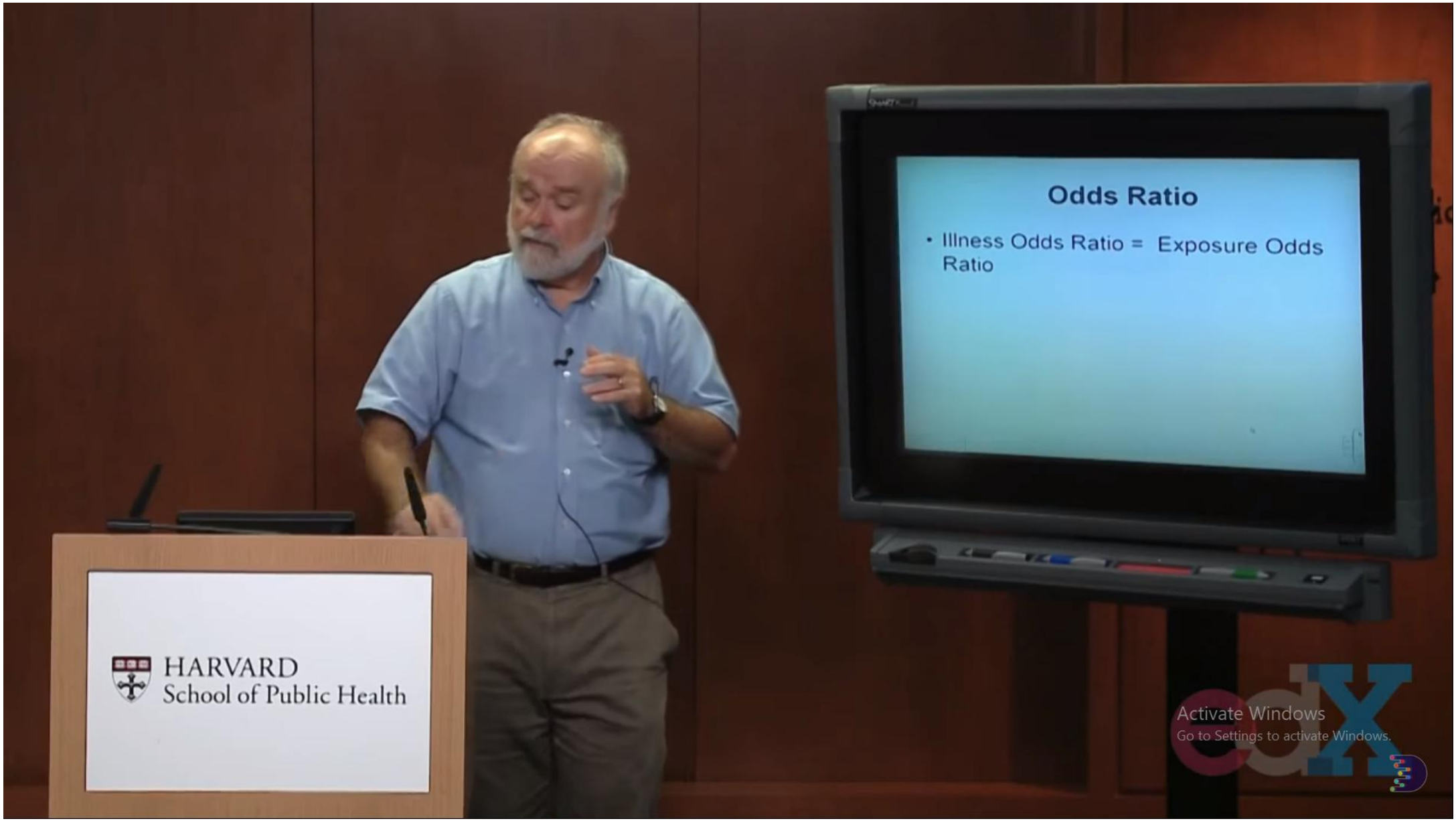
Odds Ratio

- **Odds** for not developing headache = 1/(odds for developing disease)

$$40/10 = 1/(10/40)$$

- **Odds Ratio** for not developing outcome = 1/(Odds Ratio for developing outcome)

$$(40/10)/(45/5) = .4444 = 1/2.25$$


Odds Ratio
• Illness Odds Ratio = Exposure Odds Ratio
HARVARD
School of Public Health
Activate Windows
Go to Settings to activate Windows.

	Headache	No Headache	**Disease Odds**
Males	10	40	**10/40**
Females	5	45	**5/45**
Exposure Odds			

	Headache	No Headache	**Disease Odds**
Males	10	40	**10/40**
Females	5	45	**5/45**
Exposure Odds	10/5		

	Headache	No Headache	**Disease Odds**
Males	10	40	**10/40**
Females	5	45	**5/45**
Exposure Odds	10/5	40/45	

	Headache	No Headache	Disease Odds
Males	10	40	**10/40**
Females	5	45	**5/45**
Exposure Odds	10/5	40/45	

EOR = (10/5) /

	Headache	No Headache	**Disease Odds**
Males	10	40	**10/40**
Females	5	45	**5/45**
Exposure Odds	10/5	40/45	

$$EOR = (10/5) / (40/45) = 2.25$$

Odds Ratio

- Illness Odds Ratio = Exposure Odds Ratio

$$(10/40)/(4/45) = (10/5)/(40/45)$$

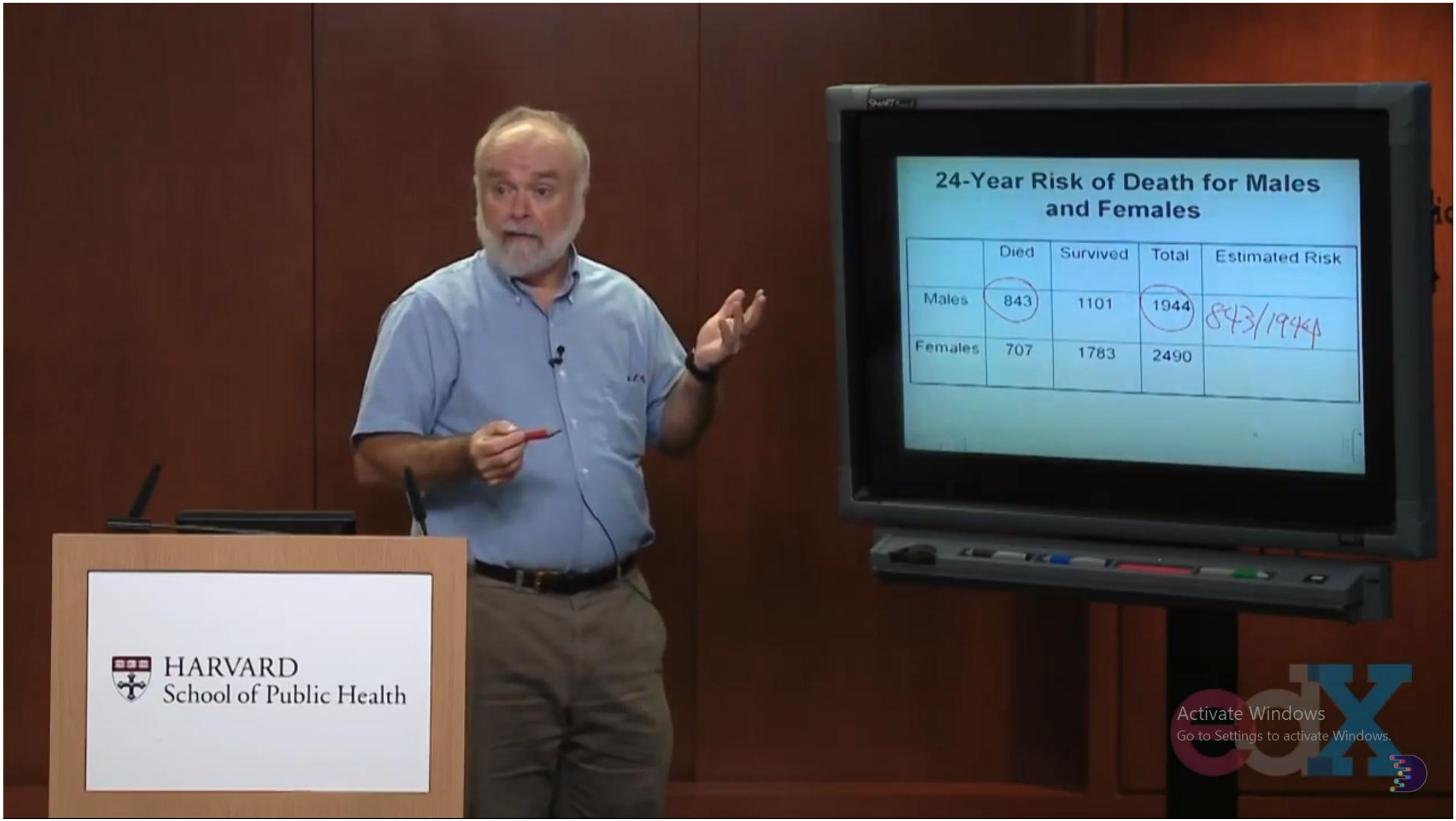
- For small risks, OR≈ RR

Risk of Death by Sex in Framingham Heart Study Data

	Died	Survived	Total
Males	843	1101	1944
Females	707	1783	2490

Questions

- What is the 24-year Risk Ratio for dying, comparing Males (exposed group) to Females (non-exposed group)?
- What is the 24-year Risk Difference for dying, comparing Males (exposed group) to Females (non-exposed group)?
- What is the 24-year Odds Ratio for dying, comparing Males (exposed group) to Females (non-exposed group)?


24-Year Risk of Death for Males and Females
Died
Survived
Total
Estimated Risk
Males
843
1101
1944
843/1944
Females
707
1783
2490
HARVARD
School of Public Health
Activate Windows
Go to Settings to activate Windows.

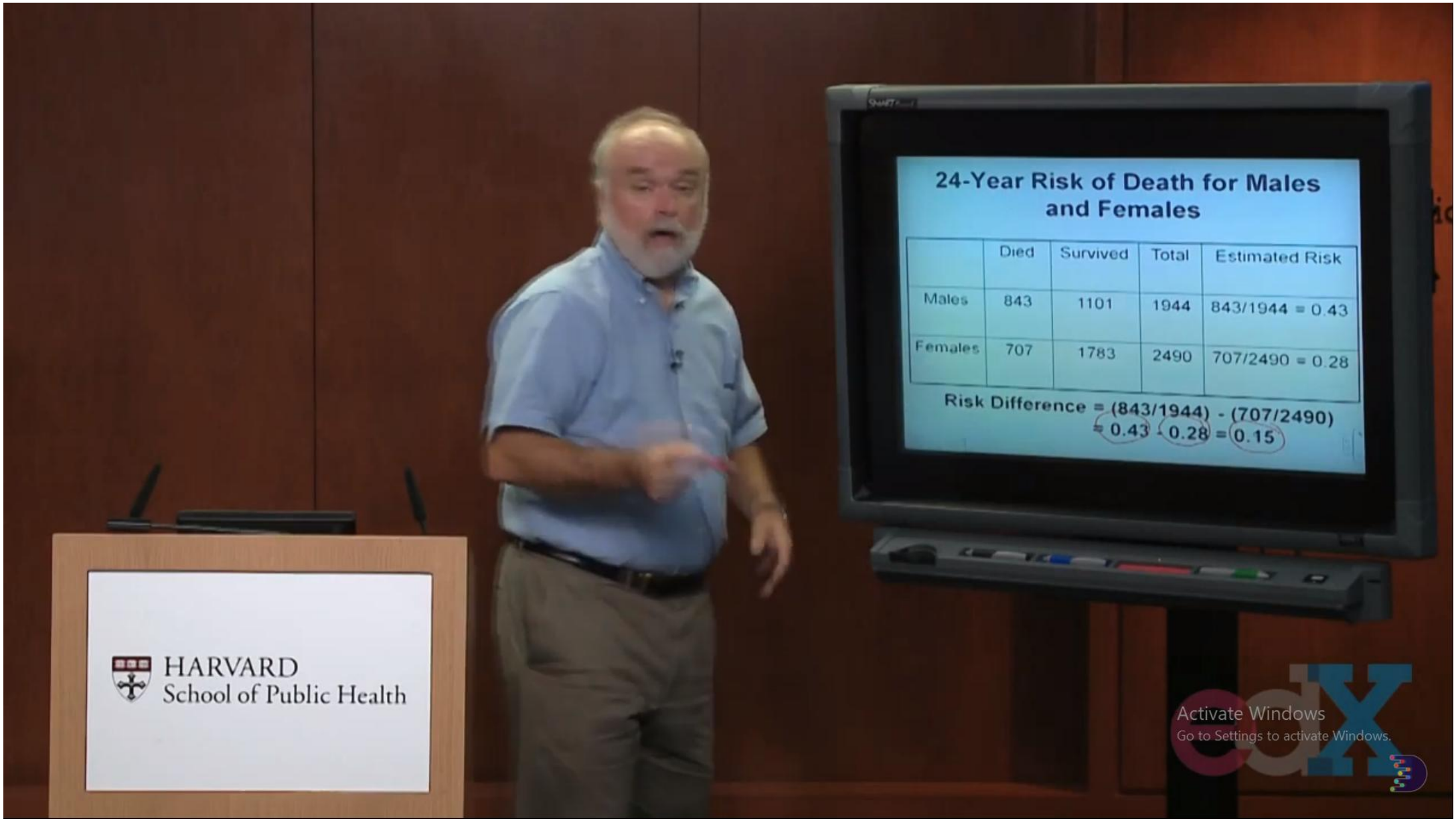
24-Year Risk of Death for Males and Females

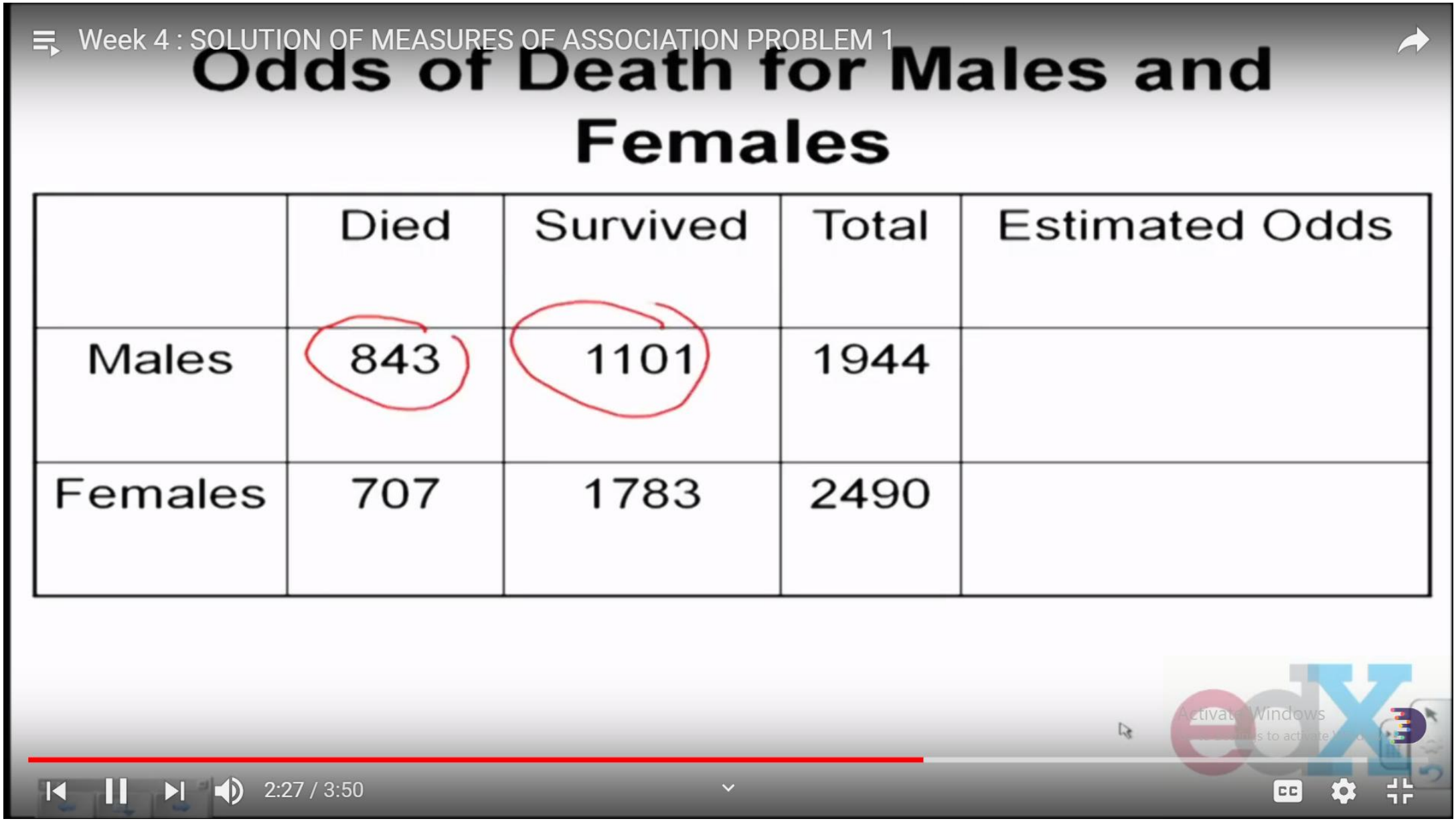
	Died	Survived	Total	Estimated Risk
Males	843	1101	1944	843/1944
Females	707	1783	2490	

24-Year Risk of Death for Males and Females

	Died	Survived	Total	Estimated Risk
Males	843	1101	1944	843/1944 = 0.43
Females	707	1783	2490	707/2490 = 0.28

Risk Ratio = (843/1944) / (707/2490)
= 0.43 / 0.28 = 1.54


24-Year Risk of Death for Males and Females
Died
Survived
Total
Estimated Risk
Males
843
1101
1944
843/1944 = 0.43
Females
707
1783
2490
707/2490 = 0.28
Risk Difference = (843/1944) - (707/2490)
= 0.43 - 0.28 = 0.15
HARVARD
School of Public Health
Activate Windows
Go to Settings to activate Windows.


Week 4 : SOLUTION OF MEASURES OF ASSOCIATION PROBLEM 1
Odds of Death for Males and Females
Died
Survived
Total
Estimated Odds
Males
843
1101
1944
Females
707
1783
2490
Activate Windows
2:27 / 3:50

Odds of Death by Sex in Framingham Heart Study Data

	Died	Survived	Total	Estimated Odds
Males	843	1101	1944	843/1101 = 0.77
Females	707	1783	2490	707/1783 = 0.40

Odds Ratio = (843/1101) / (707/1782)
= 0.77 / 0.40 = 1.93

Rate of Death by Sex in Framingham Heart Study Data

	Died	Person-Years of Follow-up
Males	843	38,287.33
Females	707	52,828.15

Rate of Death by Sex in Framingham Heart Study Data

	Died	Person-Years of Follow-up
Males	843	38,287.33
Females	707	52,828.15

Rate of Death for Males and Females

	Died	Person-Years of Follow-up	Incidence Rate
Males	843	38,287.33	843/38,287.33 py
Females	707	52,828.15	

Rate of Death for Males and Females

	Died	Person-Years of Follow-up	Incidence Rate
Males	843	38,287.33	843/38,287.33py = 2.20/100py
Females	707	52,828.15	707/52,828.15py = 1.34/100py

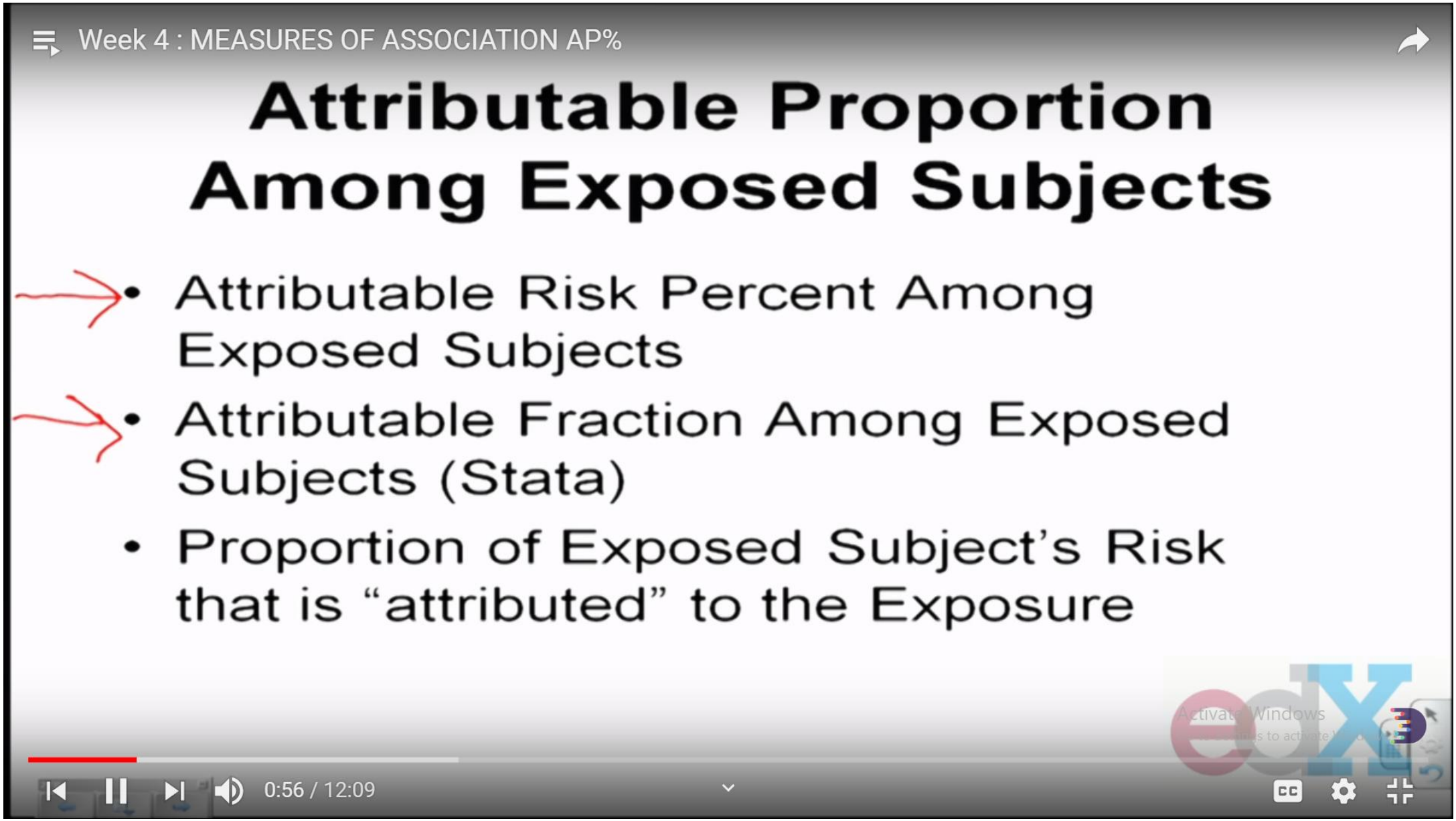
Rate of Death for Males and Females

	Died	Person-Years of Follow-up	Incidence Rate
Males	843	38,287.33	843/38,287.33py = 2.20/100py
Females	707	52,828.15	707/52,828.15py = 1.34/100py

Rate Ratio = (843/38287.33py) / (707/52828.16py)
= 1.65

Attributable Proportion Among Exposed Subjects

- Attributable Risk Percent Among Exposed Subjects
- Attributable Fraction Among Exposed Subjects (Stata)
- Proportion of Exposed Subject's Risk that is "attributed" to the Exposure


Week 4 : MEASURES OF ASSOCIATION AP%
Attributable Proportion Among Exposed Subjects
Attributable Risk Percent Among Exposed Subjects
Attributable Fraction Among Exposed Subjects (Stata)
Proportion of Exposed Subject's Risk that is "attributed" to the Exposure
0:56 / 12:09

Attributable Proportion Among Exposed Subjects

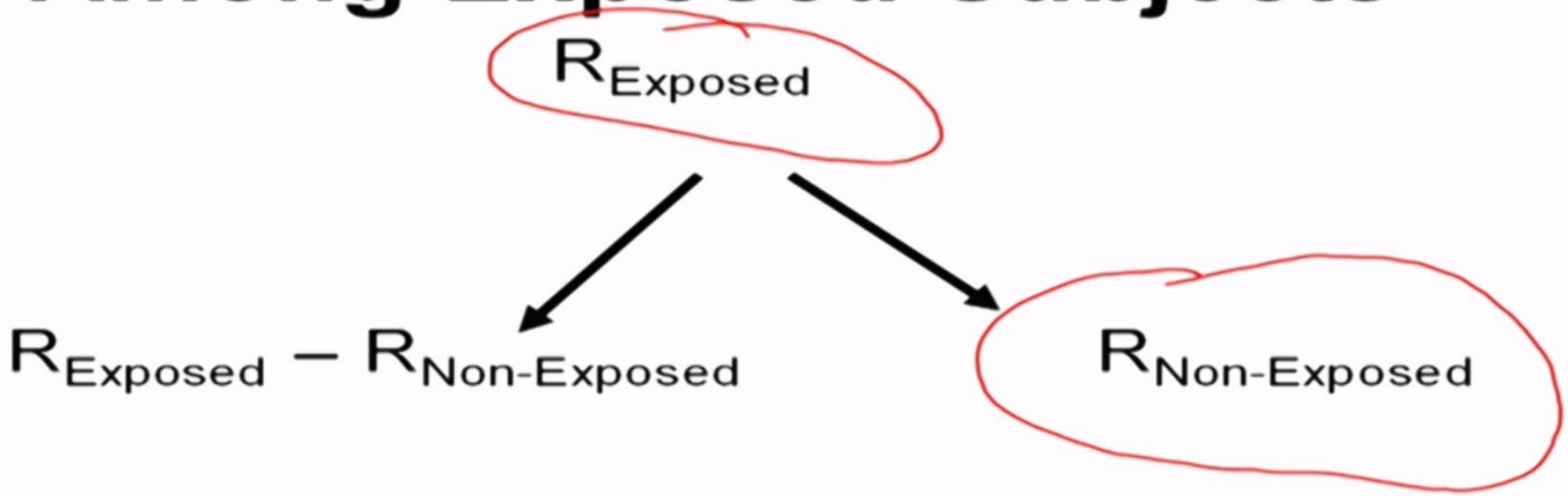
$$R_{Exposed}$$

$$R_{Exposed} - R_{Non\text{-}Exposed} \qquad R_{Non\text{-}Exposed}$$

$$AP\%_{Exposed} = (R_{Exposed} - R_{Non\text{-}Exposed})/R_{Exposed}$$

$$= (RR - 1)/RR$$

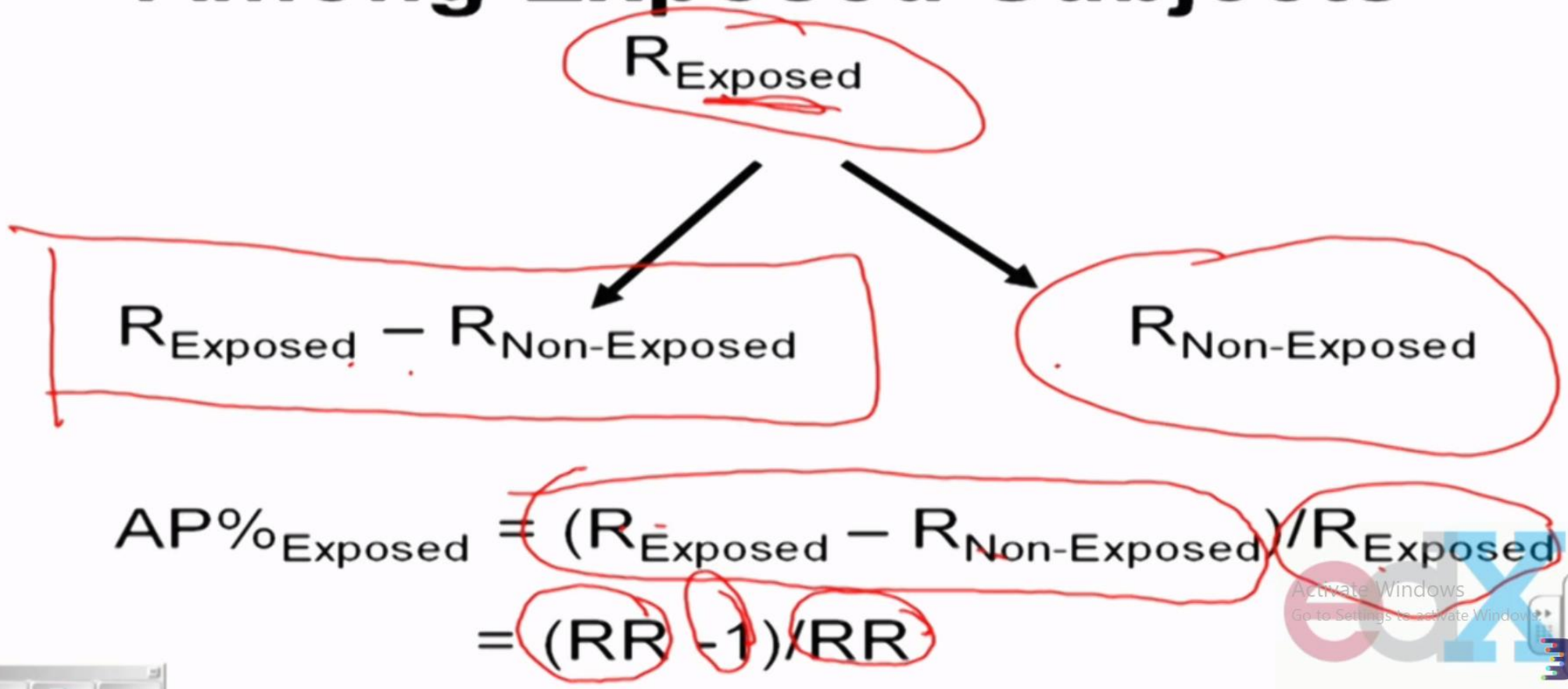
Attributable Proportion Among Exposed Subjects



$$AP\%_{Exposed} = (R_{Exposed} - R_{Non\text{-}Exposed})/R_{Exposed}$$

$$= (RR - 1)/RR$$

Attributable Proportion Among Exposed Subjects



$$AP\%_{Exposed} = (R_{Exposed} - R_{Non\text{-}Exposed})/R_{Exposed}$$

$$= (RR - 1)/RR$$

24-Year Risk of Death Among Smokers and Non-Smokers

	Died	Survived	Total	Estimated Risk
Smokers	788	1393	2181	788/2181 = 0.36
Non-Smokers	762	1491	2253	762/2253 = 0.34
Total	1550	2884	4434	1550/4434 = 0.35

RR = (788/2181) / (762/2253) = 1.07

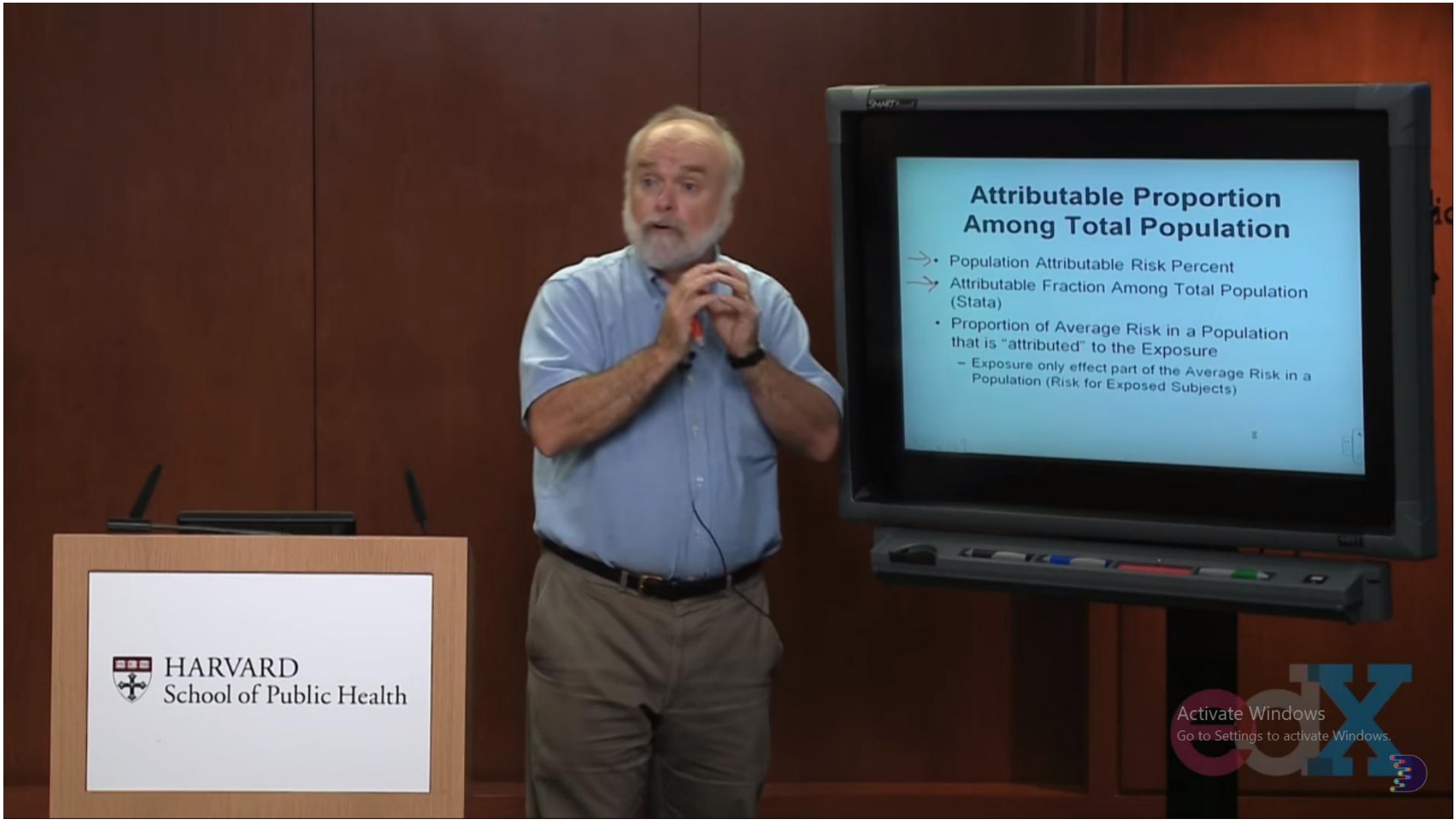
$AP\%_{Exposed}$ = (1.07 -1) /1.07 = 0.065

24-Year Risk of Death Among Smokers and Non-Smokers

	Died	Survived	Total	Estimated Risk
Smokers	788	1393	2181	788/2181 = 0.36
Non-Smokers	762	1491	2253	762/2253 = 0.34
Total	1550	2884	4434	1550/4434 = 0.35

RR = (788/2181) / (762/2253) = 1.07

$AP\%_{Exposed}$ = (1.07 -1) /1.07 = 0.065


Attributable Proportion
Among Total Population
Population Attributable Risk Percent
Attributable Fraction Among Total Population (Stata)
Proportion of Average Risk in a Population that is "attributed" to the Exposure
Exposure only effect part of the Average Risk in a Population (Risk for Exposed Subjects)
HARVARD
School of Public Health
Activate Windows
Go to Settings to activate Windows.

Attributable Proportion Among Total Population

$$R_{Average}$$

$$R_{Average} - R_{Non\text{-}Exposed} \qquad R_{Non\text{-}Exposed}$$

$$AP\%_{Population} = (R_{Average} - R_{Non\text{-}Exposed})/R_{Average}$$

Attributable Proportion Among Total Population

$$R_{Average} = (p)(R_{Exposed}) + (1-p)R_{Non-Exposed}$$

where p = Prevalence of Exposure

$$AP\%_{Population} = (R_{Average} - R_{Non-Exposed})/R_{Average}$$

$$= \frac{[(p)R_{Exposed} + (1-p)R_{Non-Exposed}] - R_{Non-Exposed}}{(p)R_{Exposed} + (1-p)R_{Non-Exposed}}$$

$$= p(RR-1)/(1 + p(RR-1))]$$

Attributable Proportion Among Total Population

$$R_{Average} = (p)(R_{Exposed}) + (1-p)R_{Non\text{-}Exposed}$$

where p = Prevalence of Exposure

$$AP\%_{Population} = (R_{Average} - R_{Non\text{-}Exposed})/R_{Average}$$

$$= \frac{[(p)R_{Exposed} + (1-p)R_{Non-Exposed}] - R_{Non-Exposed}}{(p)R_{Exposed} + (1-p)R_{Non-Exposed}}$$

$$= p(RR-1)/(1 + p(RR-1))$$

24-Year Risk of Death Among Smokers and Non-Smokers

	Died	Survived	Total	Estimated Risk
Smokers	788	1393	2181	788/2181= 0.36
Non-Smokers	762	1491	2253	762/2253 = 0.34
Total	1550	2884	4434	1550/4434 = 0.35

$p = P(\text{Smokers}) = 2181/4434 = .49$

$AP\%_{Population} = .49(1.07 - 1) / [1 + .49(1.07 - 1)] = 0.033$

24-Year Risk of Death Among Smokers and Non-Smokers

	Died	Survived	Total	Estimated Risk
Smokers	788	1393	2181	788/2181= 0.36
Non-Smokers	762	1491	2253	762/2253 = 0.34
Total	1550	2884	4434	1550/4434 = 0.35

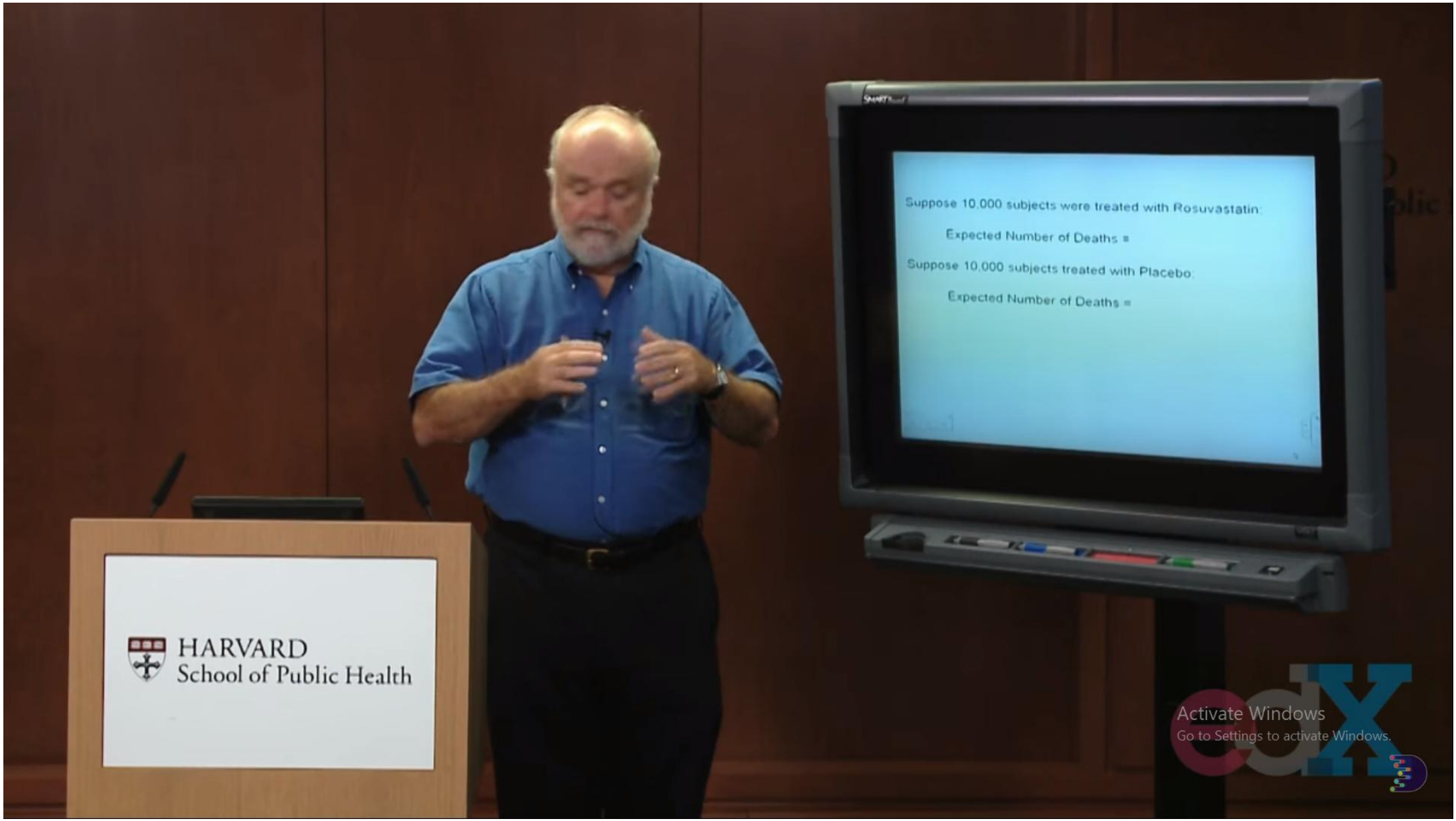
p = P(Smokers) = 2181/4434 = .49

$AP\%_{Population}$ = .49(1.07 -1) /[1+ .49(1.07-1)] = 0.033

Example: Rosuvastatin to Prevent Mortality in Subjects with Elevated C-Reactive Protein

Treatment	Deaths	Number of Subjects	Estimated Risk
Rosuvastatin	198	8901	0.0222
Placebo	247	8901	0.0277

Ridker et al. N Engl J Med 2008;359:2195-207



Suppose 10,000 subjects were treated with Rosuvastatin:

Expected Number of Deaths = 10,000 × 0.0224

Suppose 10,000 subjects treated with Placebo:

Expected Number of Deaths =

Suppose 10,000 subjects were treated with Rosuvastatin:

Expected Number of Deaths = 10,000 x .0222 = 222

Suppose 10,000 subjects treated with Placebo:

Expected Number of Deaths =

Suppose 10,000 subjects were treated with Rosuvastatin:

Expected Number of Deaths = 10,000 x .0222 = 222

Suppose 10,000 subjects treated with Placebo:

Expected Number of Deaths = 10,000 x .0277 = 277

Suppose 10,000 subjects were treated with Rosuvastatin:

Expected Number of Deaths = 10,000 x .0222 = 222

Suppose 10,000 subjects treated with Placebo:

Expected Number of Deaths = 10,000 x .0277 = 277

Implications:

277 – 222 = **55** deaths prevented for every **10,000** treated subjects

Suppose 10,000 subjects were treated with Rosuvastatin:

Expected Number of Deaths = 10,000 x .0222 = 222

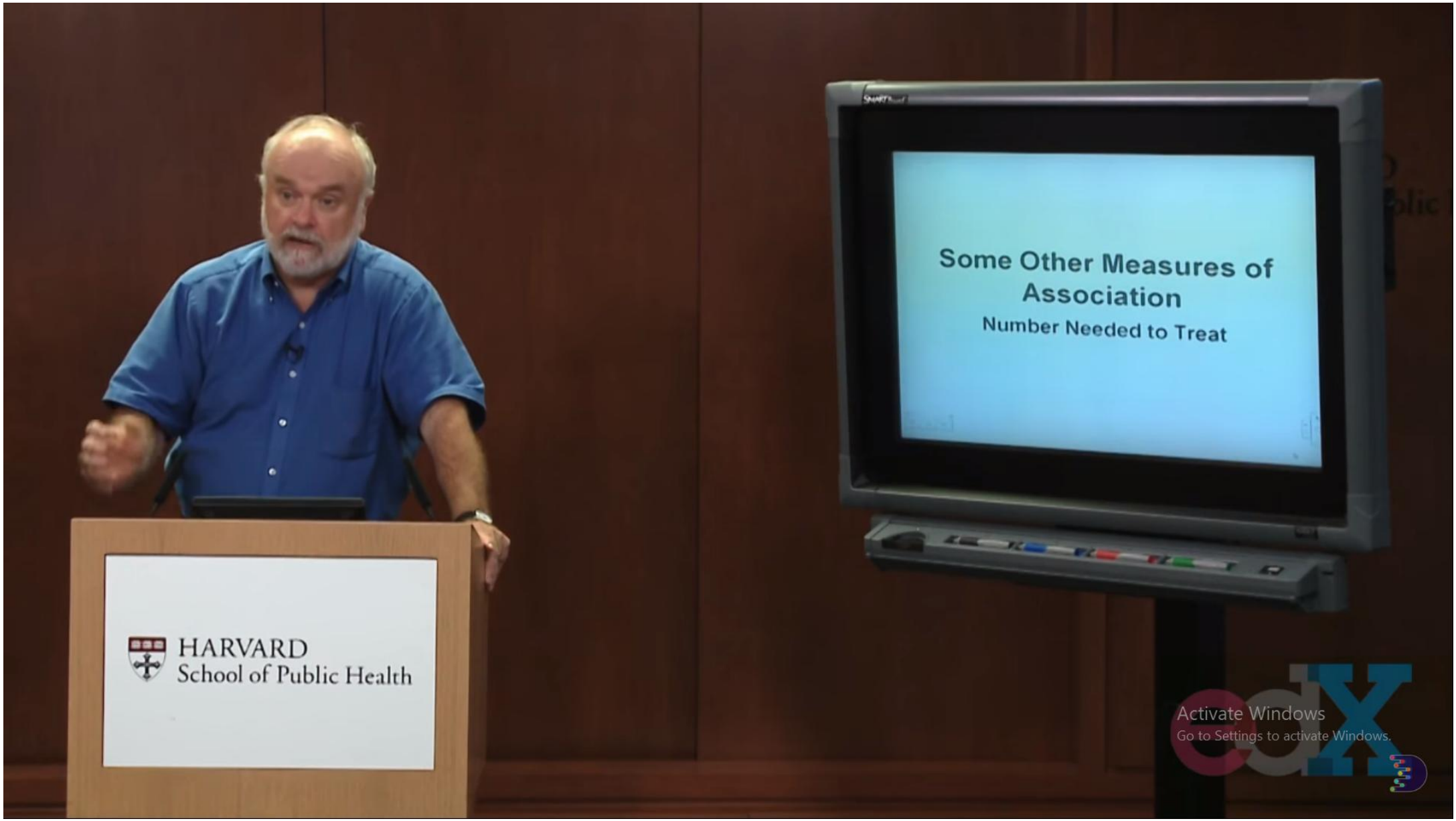
Suppose 10,000 subjects treated with Placebo:

Expected Number of Deaths = 10,000 x .0277 = 277

Implications:

277 – 222 = **55** deaths prevented for every **10,000** treated subjects

→ **1** (55/55) death prevented for every **181.8 (10,000/55)** treated subjects


Some Other Measures of Association
Number Needed to Treat
HARVARD
School of Public Health

Number Needed to Treat (NNT)

NNT= number of subjects needed to treat to prevent 1 outcome

Example:

181.8 treated with Rosuvastatin would prevent 1 death

$$181.8 = 1/(Risk_{Placebo} - Risk_{Treatment}) = 1/(\text{Risk Difference})$$

$$= 1/(0.0277 - 0.0222)$$
$$= 1/(0.0055)$$
$$= 1/(55/10{,}000) = 10{,}000/55 = 181.8$$

group and then dividing that quantity into the number one this

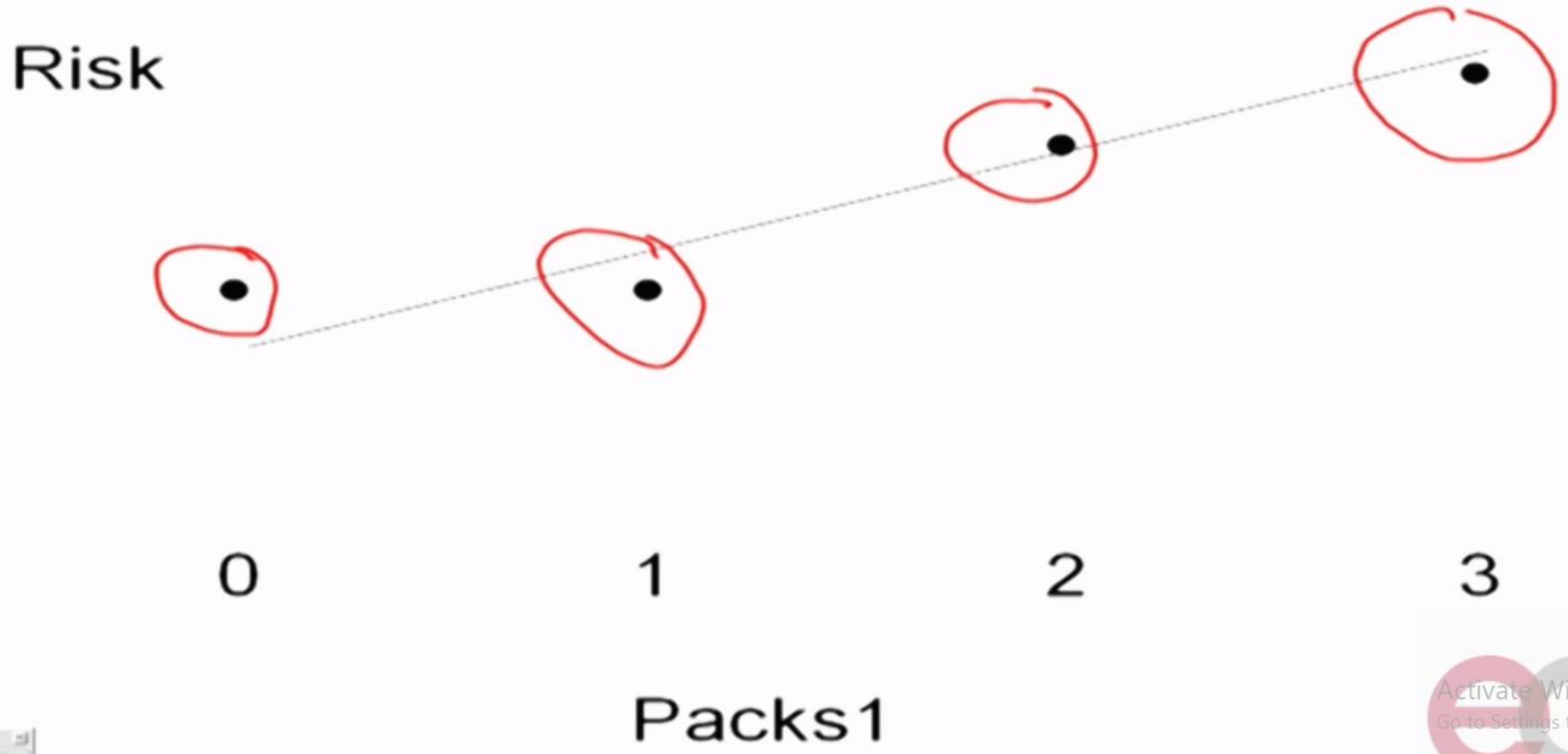
24-Year Risk of Death by Packs of Cigarettes Smoked (FHS)

Packs1	Died	Survived	Total	Estimated Risk
0	762	1491	2253	762/2253 = .34
1	573	1098	1671	573/1671 = .34
2	169	229	398	169/398 = .43
3	36	44	80	36/80 = .45

24-Year Risk of Death by Packs of Cigarettes Smoked (FHS)

Packs1	Died	Survived	Total	Estimated Risk
0	762	1491	2253	762/2253 = .34
1	573	1098	1671	573/1671 = .34
2	169	229	398	169/398 = .43
3	36	44	80	36/80 = .45

24-Year Risk of Death by Packs of Cigarettes Smoked (FHS)



Equation of a Straight Line: $Y = mX + b$

$$\text{Slope} = m = \Delta Y/(\text{unit } \Delta X)$$

Regression Model Equation:

$$\text{Risk} = B_0 + B_1(\#\ \text{Packs1})$$

Slope (B_1) : Estimate of the Effect of Smoking

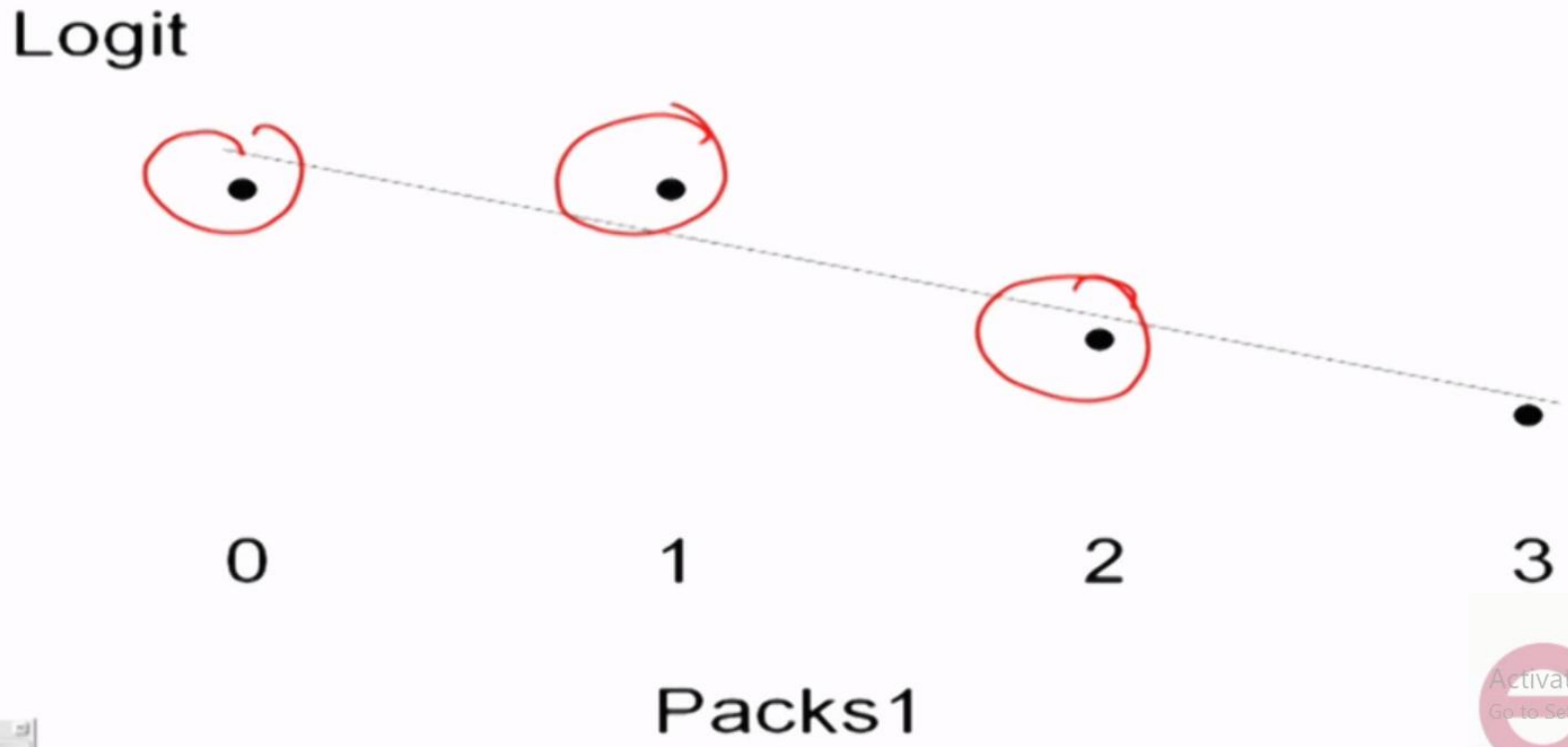
$$B_1 = (\Delta \text{ in Risk})/(\text{Smoking 1 additional pack})$$

24-Year Risk of Death by Packs of Cigarettes Smoked (FHS)

Packs1	Died	Survived	Total	Log(Odds)* (Logit)
0	762	1491	2253	Log(762/1491) = -.67
1	573	1098	1671	Log(573/1098) = -.65
2	169	229	398	Log(169/229) = -.30
3	36	44	80	Log(36/44) = -.20

* Log(Odds) = ln (Odds) = natural log of the Odds of Death

24-Year Risk of Death by Packs of Cigarettes Smoked (FHS)



Logistic Regression Model Equation:

$$\text{Log(Odds)} = B_0 + B_1(\#\ \text{Packs1})$$

Slope (B_1) : Estimate of the Effect of Smoking

B_1 = (Δ in Log(Odds))/(Smoking 1 additional pack)

Some Math:

$$\log(P/(1-P)) = B_0 + B_1(\text{Packs})$$

$$\text{Packs1}=0 : \log(P_0/(1-P_0)) = B_0$$
$$\text{Packs1}=1 : \log(P_1/(1-P_1)) = B_0 + 1B_1$$
$$\text{Packs1}=2 : \log(P_2/(1-P_2)) = B_0 + 2B_1$$
$$\text{Packs1}=3 : \log(P_3/(1-P_3)) = B_0 + 3B_1$$

$$B_1 = [\log(P_{x+1}/(1-P_{X+1}))] - [\log(P_x/(1-P_X))]$$
$$= \log([P_{x+1}/(1-P_{X+1})] / [(P_x/(1-P_X)])$$
$$= \log(\textbf{Odds Ratio})$$

Recall: Laws of Logarithms

$$\log(A) - \log(B) = \log\,(A/B)$$

Implications

- Regression coefficients = slopes
- Measures of Association
- Regression coefficient from logistic regression model is a log(Odds Ratio)