Week 2: Stratified Analysis & Effect Measure Modification

Video 1: Bias and Efficiency

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

Key Concepts

- Definitions of bias and efficiency
- Notation
- Hypothesis tests
- Point and interval estimates for the rate ratio
- Point and interval estimates for the rate difference

Trade-off Between Bias and Efficiency

- In general, there is a trade-off between **bias** and **efficiency**

- In the most extreme case, we can stratify so finely and on so many risk factors that there are very few data in any one stratum to serve as a basis for the comparison of the exposed to the unexposed, or cases to controls. Thus, the **efficiency** with which we estimate the effect is severely reduced

- However, since we control so carefully for all confounders, there should be no residual **bias** due to confounding

Definition of Bias

- The difference between the expected value of an estimation procedure and the true value that the procedure is attempting to estimate. If this difference is zero, the estimate is unbiased.

- For example, given sufficient data, the $\hat{RR}_{MH}$ is an unbiased estimate of the true RR under the assumption of no effect measure modification, no residual confounding, no confounding by unmeasured risk factors, no selection bias and no information bias. Mathematically we write:

$$\text{BIAS} = 0 \quad \text{if} \quad E(\hat{RR}_{MH}) - RR = 0,$$

$$\text{where } E(\hat{RR}_{MH}) = \int_{0}^{+\infty} \hat{RR}_{MH} \Pr(\hat{RR}_{MH} \mid \text{data}) \, d\hat{RR}_{MH}$$

Bias Versus Efficiency

- **Validity**: lack of bias. Only unbiased procedures are *valid*.

- **Efficiency**: the precision of an estimation procedure
  - The smaller the variance of a procedure, the more efficient it is.
  - We may collapse sparse strata, trading a reduction in the validity of our estimate due to residual confounding for an improvement in its efficiency

BREAK

Week 2: Stratified Analysis & Effect Measure Modification

Video 2: Notation for Stratified Person-time Data

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

Key Concepts

- Definitions of bias and efficiency
- Notation
- Hypothesis tests
- Point and interval estimates for the rate ratio
- Point and interval estimates for the rate difference

Notation for Person-time Data
(open cohort and closed cohort studies)

- Recall our notation for an unstratified table of person-time data in open cohort and closed cohort studies:

Person-Time Data

	E	$\bar{E}$	
Cases	**a**	**b**	M_1
Person-Time	N_1	N_0	T

Notation for Person-time Data
(open cohort and closed cohort studies)

- Now we stratify the data by one or more confounding variables, so that each stratum consists of subjects who have, on average, the same risk for disease, with the possible exception of the exposure effect.

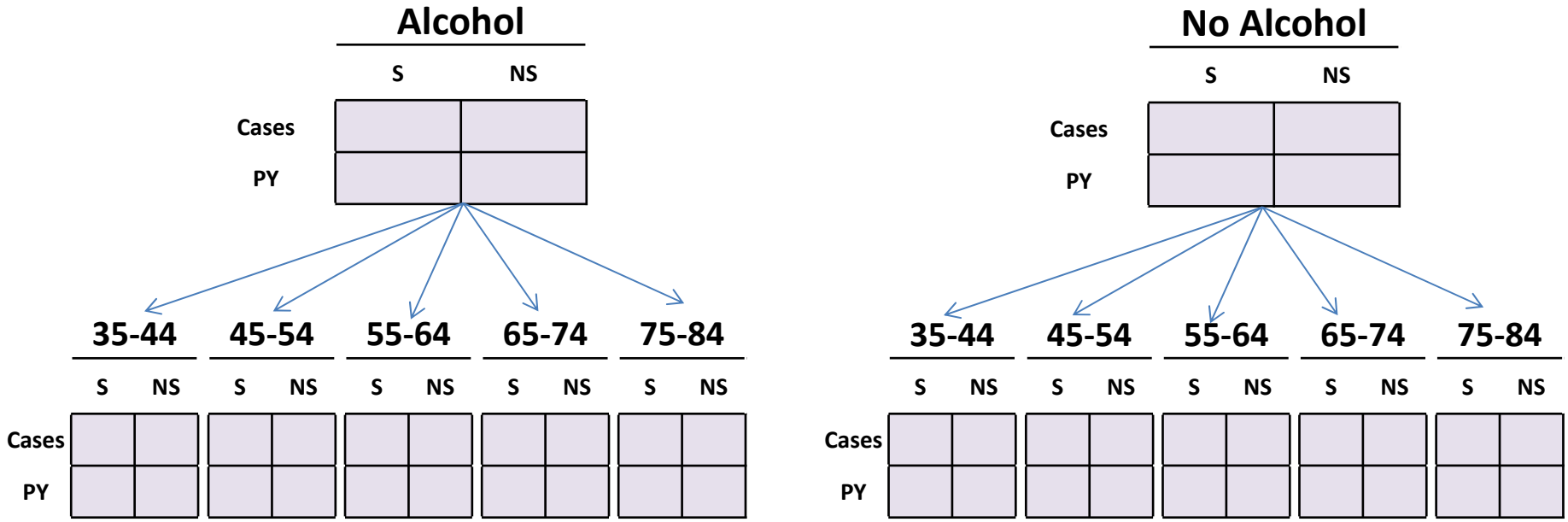
- We have i=1,...,I of these strata, which are formed by each unique combination of levels of the confounding variables for which there are data.

Stratified Person-Time Data

	E	$\bar{E}$	
Cases	a_i	b_i	M_{1i}
Person-Time	N_{1i}	N_{0i}	T_i

Stratification on Several Confounders

- If the risk of death from CHD increases with age, and the prevalence of smoking changes with age, we will want to control for age (**5** strata).
- In addition, the prevalence of alcohol drinking may also vary with smoking status, and moderate alcohol drinking may be protective for the development of CHD -- thus, we may wish to control for alcohol drinking as well as age.
- In this case, there will be **10** strata:



British Doctors Study Revisited

- We are investigating the association between cigarette smoking and coronary heart disease (CHD) mortality
- As seen previously, the crude table, summing over the appropriate cells, is:

	S	NS
Cases	630	101
Person-years	142,247	39,220
IR Per 10^4 PY	44.29	25.75

- We stratify by age (in decades) to form five age strata
- Smoking: S=smoker, NS=non-smoker

Age	**35-44**		**45-54**		**55-64**		**65-74**		**75-84**	
	S	**NS**	**S**	**NS**	**S**	**NS**	**S**	**NS**	**S**	**NS**
Cases	32	2	104	12	206	28	186	28	102	31
PY	52,407	18,790	43,248	10,673	28,612	5,710	12,663	2,585	5,317	1,462
IR Per 10^4 PY	6.11	1.06	24.1	11.2	72.0	49.0	146.9	108.3	191.8	212.0

BREAK

Week 2: Stratified Analysis & Effect Measure Modification

Video 3: Hypothesis Testing with Stratified Data

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

Key Concepts

- Definitions of bias and efficiency
- Notation
- Hypothesis tests
- Point and interval estimates for the rate ratio
- Point and interval estimates for the rate difference

Hypothesis Test for Unstratified Data

- Recall the hypothesis test statistic:

$$Z^2 = \frac{[X - E(X|H_0)]^2}{Var(X|H_0)}$$

- In open cohorts with no confounding,
  - X = number of exposed cases = a
  - $E(X|H_o)$ = number of exposed cases expected under H_o
    = total number of cases * $Pr(E) = M_1(N_1/T)$
  - $Var(X|H_0) = M_1 \frac{N_1}{T}(1 - \frac{N_1}{T}) = \frac{M_1 N_1 N_0}{T^2}$

Hypothesis Test for Stratified Data (1)

- In open cohorts with confounding, we stratify the data on all confounding variables to form I strata
- We then calculate the test statistic:

$$Z^2 = \frac{\left[\sum_{i=1}^{I} X_i - \sum_{i=1}^{I} E_i\left(X_i | H_0\right)\right]^2}{\sum_{i=1}^{I} Var_i\left(X_i | H_0\right)} \sim \chi_1^2$$

Hypothesis Test for Stratified Data (2)

- The summations are over each of the i=1,...,I strata.
- Each of the three components of the test statistic has the same stratum-specific form as in the crude analysis:
  - X_i = number of exposed cases in stratum i = a_i
  - $E_i(X_i | H_0)$ = number of exposed cases expected under H_0 in stratum i

$$= M_{1i} \frac{N_{1i}}{T_i}$$

  - $Var_i(X_i | H_0) = M_{1i} \frac{N_{1i}}{T_i} (1 - \frac{N_{1i}}{T_i}) = \frac{M_{1i} N_{1i} N_{0i}}{T_i^2}$
  - and Z^2 is distributed χ^2 with one degree of freedom

Hypothesis Test for Stratified Data
Null and Alternative Hypothesis

- H_0: There is no association between smoking and CHD mortality after stratifying by age.
  - $H_0: IRR_{MH} = 1 \leftrightarrow IRD_{Summ} = 0 \leftrightarrow \ln(IRR_{MH}) = 0$
- H_A: There is an association between smoking and CHD mortality after stratifying by age.
  - $H_A: IRR_{MH} \neq 1 \leftrightarrow IRD_{Summ} \neq 0 \leftrightarrow \ln(IRR_{MH}) \neq 0$

British Doctors Study
Test Statistic (1)

$$Z^2 = \frac{\left[\sum_{i=1}^{I} X_i - \sum_{i=1}^{I} E_i\left(X_i | H_0\right)\right]^2}{\sum_{i=1}^{I} Var_i\left(X_i | H_0\right)} \sim \chi_1^2$$

$$\sum X_i = \sum a_i = 630$$

$$\sum E_i\left(X_i | H_0\right) = \sum\left(\frac{M_{1i}\, N_{1i}}{T_i}\right)$$

$$= \frac{34*52407}{71197} + \frac{116*43248}{53921} + \frac{234*28612}{34322} + \frac{214*12663}{15248} + \frac{133*5317}{6779}$$

$$= 25.03 + 93.04 + 195.07 + 177.72 + 104.32$$

$$= 595.17$$

British Doctors Study
Test Statistic (2)

$$\Sigma \, Var_i(X_i \mid H_0) = \sum \left(\frac{M_{1i} N_{1i} N_{0i}}{T_i^2} \right)$$

$$= \frac{34 * 52407 * 18790}{71197^2} + \frac{116 * 43248 * 10673}{53921^2} + \frac{234 * 28612 * 5710}{34322^2}$$

$$+ \frac{214 * 12663 * 2585}{15248^2} + \frac{133 * 5317 * 1462}{6779^2}$$

$$= 110.10$$

- Thus, the test statistic is:

$$Z^2 = \frac{[\sum_{i=1}^{I} X_i - \sum_{i=1}^{I} E_i(X_i \mid H_0)]^2}{\sum_{i=1}^{I} Var_i(X_i \mid H_0)}$$

$$= \frac{[630 - 595.17]^2}{110.10} = 11.02$$

British Doctors Study
P-Value

- $Pr[\chi_1^2 > 11.02] = 0.001$

- These data are not very compatible with the state of nature described by the null. If the null were true, we would expect to observe results this extreme or more extreme 0.1% of the time (1 in a thousand iterations of this study).

- If we were interested in a testing framework, with a pre-specified 2-sided alpha of 0.05, we would reject the null hypothesis and conclude that after conditioning on age there is a statistically significant association between cigarette smoking and CHD mortality (assuming no residual confounding by age, no other confounding, no selection bias, no information bias or other source of bias).

- Recall that the crude test statistic was 26.23. Thus, controlling for confounding by age attenuated the statistical evidence against the null. Although we still would reject the null, the crude analysis suggested that the observed results (or more extreme) would only be expected to occur 1 time in a million iterations of the study.

BREAK

Week 2: Stratified Analysis & Effect Measure Modification

Video 4: Point and Interval Estimates for Incidence Rate Ratios

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

Key Concepts

- Definitions of bias and efficiency
- Notation
- Hypothesis tests
- Point and interval estimates for the rate ratio
- Point and interval estimates for the rate difference

Incidence Rate Ratio
Point and Interval Estimates

- To calculate the estimate of the summary rate ratio, we use a weighted sum of the stratum-specific estimates:

$$\hat{IRR} = \frac{\sum_{i=1}^{I} w_i' \; \hat{IRR}_i}{\sum_{i=1}^{I} w_i'} = \sum_{i=1}^{I} w_i \, \hat{IRR}_i$$

- Where:
  - w_i' = the weight for the estimated rate ratio in stratum I
  - IRR_i = the estimated rate ratio in stratum i

$$\text{Note:} \quad w_i = \frac{w_i'}{\sum w_i'} \quad \text{and} \quad \Sigma w_i = 1$$

Choice of Weights

Stratum	35-44	45-54	55-64	65-74	75-84		
Rate Ratio	5.7	2.1	1.5	1.4	0.9		
						$\Sigma w_i'$ RR	Summary
Possible Weighting Schemes							
(a) Equal weights							
$w_i' = 1$	1	1	1	1	1	5	
$w_i = w_i'/[\Sigma w_i'] = 1/(\#\ \text{strata})$	0.20	0.20	0.20	0.20	0.20	(1)	2.32
(b) Stratum-specific person-time							
$w_i' = T_i$	71197	53921	34322	15248	6779	181467	
$w_i = w_i'/[\Sigma w_i']$ = proportion of person-time in stratum i	0.39	0.30	0.19	0.08	0.04	(1)	3.29
(c) Inverse variance weights							
$w_{i'} = \frac{1}{\hat{\mathrm{V}}\mathrm{ar}\,[\ln(\hat{\mathrm{RR}}_i)]} = \frac{1}{\frac{1}{a_i} + \frac{1}{b_i}}$	1.88	10.8	24.4	24.4	23.8	85.3	
$w_i = w_i'/[\Sigma w_i']$ = proportion of variance in stratum i	0.02	0.13	0.29	0.29	0.28	(1)	1.48
(d) Mantel-Haenszel weights							
$w_i' = \frac{b_i N_{1i}}{T_i}$	1.47	9.62	23.3	23.3	24.3	82.0	
$w_i = w_i'/[\Sigma w_i']$ = proportion of M-H weight in stratum i	0.02	0.12	0.28	0.28	0.30	(1)	1.43

Choices of Weights
Implications

- The value of the estimated summary rate ratio depends upon the chosen weights.
- All estimates, regardless of the weights, are unbiased estimates, under the assumption, which is critical for this analysis, of no effect measure modification.
- However, some strata have more data than other strata and thus estimate the stratum-specific rate ratio with more precision.
- When there is no effect measure modification, we would like to weight these strata more heavily.
- Two reasonable choices of weights are
  - the inverse of the large-strata variances and
  - the Mantel-Haenszel weights

Choices of Weights
Inverse Variance Weights

- Let weight w_i' = inverse of large-strata variances
- Recall that $\hat{\mathrm{V}}\mathrm{ar}\,(ln\,\hat{\mathrm{IRR}}) = \frac{1}{a} + \frac{1}{b}$
- Then $w_{i'} = \frac{1}{\hat{\mathrm{V}}\mathrm{ar}_{\mathrm{i}}(ln\,\hat{\mathrm{IRR}}_{\mathrm{i}})} = \frac{1}{\frac{1}{a_i} + \frac{1}{b_i}}$

- If there are no cases observed in either the exposed or unexposed group in a given stratum, the large-strata variance for that stratum becomes undefined and we lose all information from that stratum.
- In addition, if the number of cases observed in the exposed or unexposed groups within strata is small, then the large-strata variance formula may be inaccurate.

Choices of Weights
Mantel-Haenszel Weights

- Thus, to make the best use of our data, we prefer the Mantel-Haenszel weights:

$$w_i' = \frac{b_i N_{1i}}{T_i}$$

- Using the Mantel-Haenszel weights, the summary incidence rate ratio is:

$$\hat{IRR}_{MH} = \frac{\sum_{i=1}^{I} w_i' \hat{IRR}_i}{\sum_{i=1}^{I} w_i'} = \frac{\sum_{i=1}^{I} \frac{b_i N_{1i}}{T_i} \left(\frac{a_i}{N_{1i}}\right) / \left(\frac{b_i}{N_{0i}}\right)}{\sum_{i=1}^{I} \frac{b_i N_{1i}}{T_i}} = \frac{\sum_{i=1}^{I} \frac{a_i N_{0i}}{T_i}}{\sum_{i=1}^{I} \frac{b_i N_{1i}}{T_i}}$$

British Doctors Study
IRR_{MH} Estimation

- The Mantel-Haenszel summary rate ratio is:

$$\hat{IRR}_{MH} = \frac{\sum_{i=1}^{I} \frac{a_i N_{0i}}{T_i}}{\sum_{i=1}^{I} \frac{b_i N_{1i}}{T_i}}$$

$$= \frac{\frac{32 * 18790}{71197} + \frac{104 * 10673}{53921} + \frac{206 * 5710}{34322} + \frac{186 * 2585}{15248} + \frac{102 * 1462}{6779}}{\frac{2 * 52407}{71197} + \frac{12 * 43{,}248}{53921} + \frac{28 * 28612}{34322} + \frac{28 * 12663}{15248} + \frac{31 * 5317}{6779}}$$

$$= 1.425$$

- Adjusting for confounding by age attenuated the estimated rate ratio appreciably by about 17% (from 1.72 to 1.42).

- We conclude from these data that after conditioning on age, there is evidence of a 42% higher CHD death rate among smokers compared to non-smokers (assuming no residual confounding by age, no confounding by other variables, no selection bias and no information bias or other source of bias).

95% CI for $ln(\text{IRR}_{MH})$

To construct the 95% confidence interval for $ln(\text{IRR}_{MH})$, we use the usual formula:

$$ln\,\hat{\text{IRR}}_{MH} \pm 1.96\sqrt{\hat{\text{Var}}\,(ln\,\hat{\text{IRR}}_{MH})}$$

Variance of the $ln(IRR_{MH})$

- For stratified person-time data, we use the variance:

$$\hat{\mathrm{V}}\mathrm{ar}\left[\ell n(\hat{\mathrm{IRR}})\right] = \frac{\sum_{i=1}^{I} \frac{M_{1i} N_{1i} N_{0i}}{T_i^2}}{\left[\sum_{i=1}^{I} \frac{a_i N_{0i}}{T_i}\right]\left[\sum_{i=1}^{I} \frac{b_i N_{1i}}{T_i}\right]}$$

British Doctors Study
IRR_{MH} Confidence Interval Estimation

- In this example, $\hat{\mathrm{V}}\mathrm{ar}\,(\,ln\,\hat{\mathrm{I}}\mathrm{R}\mathrm{R}\,) = 0.01149$
- Thus, the 95% confidence interval for ln (IRR_{MH}) is:

$$ln\,(1.42) \pm 1.96\,\sqrt{0.01149} = (0.141, 0.560)$$

- and the 95% confidence interval for IRR_{MH} is:

$$e^{(0.141,\ 0.560)} = (1.15,\ 1.76)$$

British Doctors Study
IRR_{MH} Confidence Interval Interpretation

- Assuming no residual confounding by age, no confounding by other variables, no selection bias, no information bias and no other source of bias, we conclude that there is evidence of a higher CHD mortality rate associated with smoking, ranging in magnitude from 15% to 76% higher with 95% confidence.

- Because it has a relatively narrow range, the confidence interval indicates a good power in the data to estimate the rate ratio, and the estimate is relatively precise.

- The crude rate ratio 1.72 and confidence interval (1.39, 2.12) overestimated the impact of smoking, because of upward bias due to confounding by age. Furthermore, the CI was incorrectly centered and of incorrect width. The statistical inferences in the crude analysis are not valid.

BREAK

Week 2: Stratified Analysis & Effect Measure Modification

Video 5: Point and Interval Estimates for Incidence Rate Differences

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

Key Concepts

- Definitions of bias and efficiency
- Notation
- Hypothesis tests
- Point and interval estimates for the rate ratio
- Point and interval estimates for the rate difference

Incidence Rate Difference
Point and Interval Estimates

- To calculate the estimated summary rate difference, we use a weighted average of the stratum-specific rate differences:

$$\hat{IRD} = \frac{\sum_{i=1}^{I} w_i' \; \hat{IRD}_i}{\sum_{i=1}^{I} w_i'} = \sum_{i=1}^{I} w_i \; \hat{IRD}_i$$

- Where:
  - w_i' = the weight for the estimated rate difference in the i^{th} stratum
  - $\hat{IRD}_i$ = the estimated rate difference in the i^{th} stratum
  - and $w_i = \frac{w_i'}{\sum w_i'} \; and \; \sum w_i = 1$

Choice of Weights
Summary Incidence Rate Difference

- Two reasonable choices for weights are inverse variance weights and Mantel-Haenszel style weights.
- The inverse variance weights are very efficient when there is reasonably large amount of data within each stratum but can lead to biased estimates if there is sparse data within strata.
- Mantel-Haenszel style weights proposed by Greenland and Robins perform better, without bias when the data within strata are sparse and are provided in the roadmap.

British Doctors Study
$IRD_{summary}$ Estimation

- The summary rate difference:

$$\hat{IRD}_{summary} = \frac{\sum w_i IRD_i}{\sum w_i}, \text{where } w_i = \frac{N_{1i}N_{0i}}{T_i}$$

The computational form: $\hat{IRD}_{summary} = \frac{\sum\left(\frac{a_i N_{0i} - b_i N_{1i}}{T_i}\right)}{\sum \frac{N_{1i}N_{0i}}{T_i}}$

$$= \frac{\frac{32 * 18{,}790 - 2 * 52{,}407}{71{,}179} + \frac{104 * 10{,}673 - 12 * 43{,}248}{53{,}921} + \frac{206 * 5{,}710 - 28 * 28{,}612}{34{,}322} + \frac{186 * 2{,}585 - 28 * 12{,}663}{15{,}248} + \frac{102 * 1{,}462 - 31 * 5{,}317}{6{,}779}}{\frac{52{,}407 * 18{,}790}{71{,}197} + \frac{43{,}248 * 10{,}673}{53{,}921} + \frac{28{,}612 * 5{,}710}{34{,}322} + \frac{12{,}663 * 2{,}585}{15{,}248} + \frac{5{,}317 * 1{,}462}{6{,}779}}$$

= 0.001144 / person-year

- Assuming no confounding by additional variables, no residual confounding by age, no selection bias, information bias or any other source of bias, these data indicate an excess of 11.4 deaths from CHD per 10,000 PY among smokers compared with non-smokers.

95% CI for $IRD_{summary}$

The 95% confidence interval for the summary incidence rate difference is calculated in the usual way:

$$\mathrm{I}\hat{R}D_{summary} \pm 1.96\sqrt{var\left(\mathrm{I}\hat{R}D_{summary}\right)}$$

Where the variance of the estimated summary incidence rate difference is:

$$Var\left(I\hat{R}D_{summary}\right) = \frac{\sum\left(\frac{a_i N_{0i}^2 + b_i N_{1i}^2}{T_i^2}\right)}{\left(\sum \frac{N_{1i}N_{0i}}{T_i}\right)^2}$$

British Doctors Study
$IRD_{summary}$ CI Estimation and Interpretation

- In our example,

$$Var(\hat{IRD}_{summary}) = \frac{\frac{32 * 18{,}790^2 + 2 * 52{,}407^2}{71{,}197^2} + \frac{104 * 10{,}673^2 + 12 * 43{,}248^2}{53{,}921^2} + \frac{206 * 5{,}710^2 + 28 * 28{,}612^2}{34{,}322^2} + \frac{186 * 2{,}585^2 + 28 * 12{,}663^2}{15{,}248^2} + \frac{102 * 1{,}462^2 + 31 * 5{,}317^2}{6{,}779^2}}{\left(\frac{52{,}407 * 18{,}790}{71{,}197} + \frac{43{,}248 * 10{,}673}{53{,}921} + \frac{28{,}612 * 5{,}710}{34{,}322} + \frac{12{,}663 * 2{,}585}{15{,}248} + \frac{5{,}317 * 1{,}462}{6{,}779}\right)^2}$$

- Thus, the 95% confidence interval for IRD is:

$$0.001144 \pm 1.96\sqrt{9.574x10^{-8}}$$
$$= (5.4/10^4 PY\,, 17.5/10^4 PY)$$

- After stratifying on age, these data are consistent with an IRD ranging from 5.4 to 17.5 cases per 10^4 person years with 95% confidence, assuming no residual confounding, no confounding by other variables, no selection bias, information bias or any other source of bias.

British Doctors Study
Crude Vs. Stratified IRD Estimates

- Contrast the age-adjusted point and interval estimates for the rate difference to the corresponding crude values:

  1.85/1,000PY $(12.4/10^4$ PY, $24.6/10^4$ PY)

- The crude incidence rate difference is much larger than the age-adjusted rate difference.

- On the difference scale, there is evidence of substantial bias due to confounding by age.

BREAK

Week 2: Stratified Analysis & Effect Measure Modification

Video 6: Introduction to Effect Measure Modification Analysis

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

Key Concepts

- Definition of effect measure modification
- Impact on generalizability (transportability)
- Scale dependence of effect measure modification
- Tests of homogeneity
  - Incidence rate ratio and difference
  - Cumulative incidence ratio and difference
  - Odds ratio
- Relative Excess Risk due to Interaction (RERI)

Effect Measure Modification

- In the presence of effect measure modification, the magnitude of the association between exposure and disease varies according to the value of (across strata of) a third factor, which is called an effect modifier.
- Effect measure modification is an intrinsic phenomenon and cannot be eliminated from a study through clever design
- Effect measure modification is a finding to be reported rather than a bias to be avoided
- Synonyms: interaction, synergy, antagonism

Effect Measure Modification
Perioperative Beta-Blocker Therapy and Mortality

- In a study of 663,635 patients undergoing major non-cardiac surgery Lindenauer and colleagues reported on the effect of treatment with beta-blockers around the time of surgery.

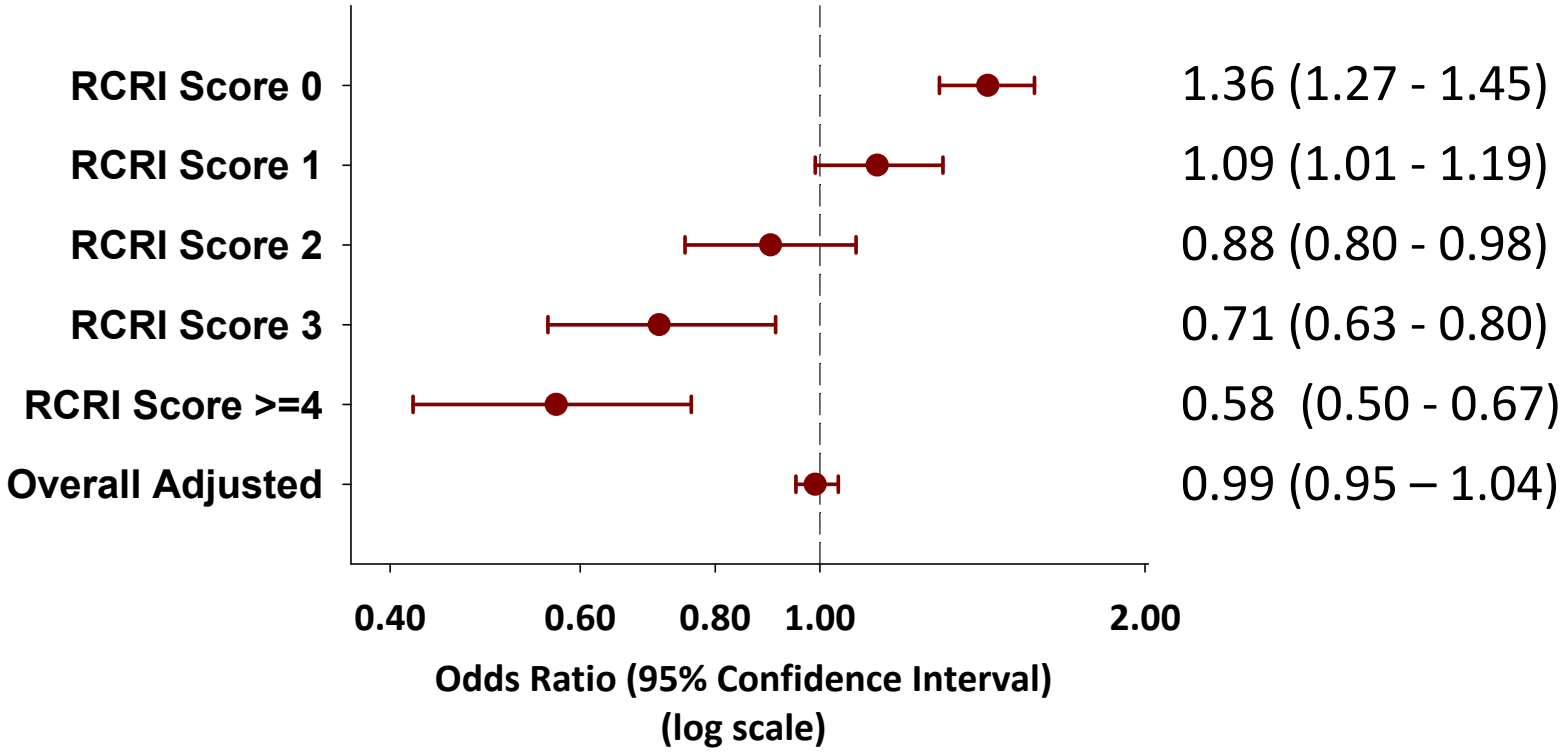
- They found that for patients who were at very low risk of cardiac complications, beta-blocker therapy was associated with increased risk of dying during the hospitalization.

- On the other hand, as the risk of cardiac complications increased, use of beta-blockers around the time of surgery became progressively more protective.

Lindenauer PK, Pekow P, Wang K, Mamidi DK, Gutierrez B, Benjamin EM. *N Engl J Med*. 2005;353(4):349-61.

Perioperative Beta-Blocker Therapy and Mortality

Perioperative beta-blocker therapy and mortality after major noncardiac surgery by Revised Cardiac Risk Index (RCRI) score



*The Revised Cardiac Risk Index is a risk assessment tool developed to predict a patient's risk of having cardiac complications associated with non-cardiac surgery; scores range from 0 (low risk) to ≥ 4 (very high risk).

Lindenauer PK, Pekow P, Wang K, Mamidi DK, Gutierrez B, Benjamin EM. *N Engl J Med*. 2005;353(4):349-61.

External Validity

- The possibility of **effect measure modification** reduces the **external validity** (generalizability or transportability) of a study which is restricted to a particular sub-population

- When the results of a study are generalizable to other, larger populations, we say that this study is **externally valid**.

BREAK

Week 2: Stratified Analysis & Effect Measure Modification

Video 7: Scale Dependence and Confounding

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

Key Concepts

- Definition of effect measure modification
- Impact on generalizability (transportability)
- Scale dependence of effect measure modification
- Tests of homogeneity
  - Incidence rate ratio and difference
  - Cumulative incidence ratio and difference
  - Odds ratio
- Relative Excess Risk due to Interaction (RERI)

Effect Measure Modification
Scale Dependence

- The presence and extent of effect measure modification depends on the scale on which the association is assessed

- When a difference measure is used, we say that the association is assessed on the additive scale

- When a ratio measure is used, we say that the association is assessed on the multiplicative scale

Effect Measure Modification
Age, Gender and Mortality Following Trigeminal Neuralgia

- Gender-related mortality in patients with trigeminal neuralgia (Rothman and Monson, 1973)

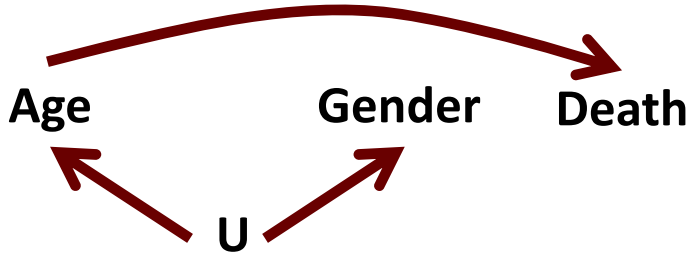
	Age < 65		Age 65+	
	Men	**Women**	**Men**	**Women**
Deaths	14	10	76	121
Person-years	1,516	1,701	949	2245
Î/10³	9.2	5.9	80.0	53.9
ÎRR	1.6		1.5	
ÎRD/10³ PY	3.3		26.2	

Scale Dependence

- On the rate ratio scale, there is no clear evidence of effect measure modification by age.
  - There is roughly a 50% higher mortality rate in the men compared with women.
  - This association is not modified by age, but remains constant in both age groups.

- On the rate difference scale, there is strong evidence of effect measure modification by age.
  - The number of excess cases among men per 1,000 person-years increases dramatically with increasing age
  - The rate difference is almost 10-times higher among those aged 65 years or more compared with the younger subjects.

Confounding vs. Effect Modification (IRR)
Age, Gender and Mortality Following Trigeminal Neuralgia



	Age < 65	Age 65+	Crude
$\hat{IRR}$	1.6	1.5	1.1

- Age is a confounder of the gender-related differential mortality among this population.

- Age is not an effect modifier on the multiplicative scale.

Confounding vs. Effect Modification (IRD)
Age, Sex and Mortality Following Trigeminal Neuralgia

	Age < 65		Age 65+	
	M	F	M	F
Deaths	14	10	76	121
Person-years	1,516	1,701	949	2,245
Î/10³	9.2	5.9	80.0	53.9
ÎRR	1.6		1.5	
ÎRD/10³ PY	3.3		26.2	

- The summary rate difference is $13.7/10^3$ PY.
- Age is a confounder of the gender-related differential mortality among this population.
- Age is an effect modifier on the additive scale.
- Due to the particular configuration of these data, the summary rate difference is about weighted slightly more heavily toward the rate difference observed among those less than 65 years of age, although most of the deaths occur in the older age group.

BREAK

Week 2: Stratified Analysis & Effect Measure Modification

Video 8: Evaluating Effect Measure Modification on the Ratio Scale

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

Key Concepts

- Definition of effect measure modification
- Impact on generalizability (transportability)
- Scale dependence of effect measure modification
- Tests of homogeneity
  - Incidence rate ratio and difference
  - Cumulative incidence ratio and difference
  - Odds ratio
- Relative Excess Risk due to Interaction (RERI)

Evaluation of Effect Measure Modification

- Effect measure modification is not the only reason for stratum-to-stratum variation in effect estimates.
- Selection bias, information bias, confounding, and chance may also produce stratum-to-stratum variations in effect estimates.
- In the case of chance, statistical tests help us assess the extent to which the differences between strata might be explained by random sampling variation.
- How do we know whether the apparent variation is real, or if it is due to random stratum-to-stratum sampling variation? We have statistical tests that will help us examine the evidence in the data to distinguish between these two possibilities.

Testing for Effect Measure Modification

We can use statistical hypothesis testing procedures to assess the data for statistical evidence of effect measure modification

Statistical hypothesis testing procedures follows a set of steps:

1. Specify the null hypothesis and its alternative.

2. Calculate a test statistic from the data which summarizes the evidence in the data in favor of, or against, the null. This test statistic must have a known probability distribution.

3. Find the p-value which corresponds to the observed value of the test statistic under the null.

4. Interpret the results.

Test of Homogeneity (IRR)
Null and Alternative Hypotheses

- H_0: The rate ratio is the same across all I levels of the stratification variable(s)
  - ↔ There is no effect modification of the IRR by the stratification variable(s)
  - ↔ The rate ratio is *homogeneous* across the strata
  - ↔ $IRR_1 = IRR_2 = \ldots = IRR_I$
  - ↔ $IRR_i = IRR_j$ for all i,j

- H_A: The rate ratio is not the same across all I levels of the stratification variable(s)
  - ↔ There is effect modification of the IRR by one (or more) of the stratification variable(s)
  - ↔ The rate ratio is *heterogeneous* across the strata
  - ↔ At least one of the IRRs does not equal at least one of the others
  - ↔ $IRR_i \neq IRR_j$ for at least one i,j pair

Test of Homogeneity (IRR)
Form of the Test Statistic

- In general, tests for homogeneity of ratio measures, i.e. for effect modification on the multiplicative scale, have the following form

$$H = \sum_{i=1}^{I} \frac{[ln(\hat{IRR}_i) - ln\,\hat{IRR}_{MH}]^2}{\hat{Var}_i[ln(\hat{IRR}_i)]} \sim \chi^2_{I-1}$$

- Where :
  - $\hat{IRR}_i$ is the stratum specific estimate of the incidence rate ratio
  - $\hat{IRR}_{MH}$ is the MH summary estimate of the incidence rate ratio, assuming homogeneity across all strata, and
  - $\hat{Var}[ln(\hat{IRR}_i)] = \frac{1}{a_i} + \frac{1}{b_i}$ is the (large-sample) variance of the stratum specific estimate of the *ln* incidence rate ratio.

ad hoc Adjustment Methods
For Zero Stratum Weight

- **Note**: If the number of exposed or unexposed cases in a stratum is 0, the weight accorded to that stratum is undefined.

- There are two *ad hoc* methods recommended to deal with this problem:
  - Add 0.5 to every cell in that stratum
  - Collapse sparse strata with similar measures of effect to eliminate 0 cells

- Neither of these options is entirely satisfactory. Why? Bias is introduced to improve efficiency.

Test of Homogeneity (IRR)
Test Statistic Computation

	Age < 65	Age 65+
$\hat{IRR}_i$	1.57	1.49
$ln(\hat{IRR}_i)$	0.452	0.396
$\hat{Var}[ln(\hat{IRR}_i)]$	0.171	0.021
IRR_{MH}	1.50	
$ln(IRR_{MH})$	0.405	

$$H = \sum_{i=1}^{I} \frac{[ln(\hat{IRR}_i) - ln\,\hat{IRR}_{MH}]^2}{\hat{Var}_i[ln(\hat{IRR}_i)]}$$

$$= \frac{(0.452-0.405)^2}{0.171} + \frac{(0.396-0.405)^2}{0.0.021} = 0.016$$

Test of Heterogeneity (IRR)
Test Statistic Interpretation

- Degrees of freedom for H = 2-1 = 1
- $Pr[\chi^2_1 > 0.016] = 0.90$
- Clearly, these stratum-specific mortality rates are homogeneous. There is no statistical evidence for effect modification by age on the multiplicative scale in these data. Of course, like all statistical tests, our interpretation depends on the validity of the assumption of freedom from bias.

BREAK

Week 2: Stratified Analysis & Effect Measure Modification

Video 9: Evaluating Effect Measure Modification on the Difference Scale

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

Key Concepts

- Definition of effect measure modification
- Impact on generalizability (transportability)
- Scale dependence of effect measure modification
- Tests of homogeneity
  - Incidence rate ratio and difference
  - Cumulative incidence ratio and difference
  - Odds ratio
- Relative Excess Risk due to Interaction (RERI)

Test of Homogeneity (IRD)
Null and Alternative Hypotheses

- H_0: The rate difference is the same across all I levels of the stratification variable(s)
  - ↔ There is no effect modification of the IRD by the stratification variable(s)
  - ↔ The rate difference is *homogeneous* across the strata
  - ↔ $IRD_1 = IRD_2 = \ldots = IRD_I$
  - ↔ $IRD_i = IRD_j$ for all i,j

- H_A: The rate difference is not the same across all I levels of the stratification variable(s)
  - ↔ There is effect modification of the IRD by one (or more) of the stratification variable(s)
  - ↔ The rate difference is *heterogeneous* across the strata
  - ↔ At least one of the IRDs does not equal at least one of the others
  - ↔ $IRD_i \neq IRD_j$ for at least one i,j pair

Test of Heterogeneity (IRD)
Form of the Test Statistic

- In general, tests for heterogeneity of difference measures, i.e. for effect modification on the additive scale, have the following form

$$H = \sum \frac{\left[I\hat{R}D_i - I\hat{R}D_{summary}\right]^2}{var(I\hat{R}D_i)}$$

- Where :
  - $I\hat{R}D_i$ is the stratum specific estimate of the incidence rate difference
  - $I\hat{R}D_{summary}$ is the summary estimate of the incidence rate difference, assuming homogeneity across all strata, and
  - $var\left(I\hat{R}D_i\right) = \frac{a_i}{N_{1i}^2} + \frac{b_i}{N_{0i}^2}$ is the variance of the stratum specific estimate of the incidence rate difference.

Test of Heterogeneity (IRD)
Test Statistic Computation

	Age < 65	Age 65+
$\hat{IRD}_i$	$3.36/10^3$PY	$26.2/10^3$PY
$\hat{Var}[IRD_i]$	$9.5/10^6$ PY^2	$108/10^6$ PY^2
$IRD_{summary}$	$13.7/10^3$ PY	

$$H = \sum \frac{\left[I\hat{R}D_i - I\hat{R}D_{summary}\right]^2}{var(I\hat{R}D_i)}$$

$$= \frac{\left(3.36/10^3 PY - 13.7/10^3 PY\right)^2}{9.55/10^6\ \mathrm{PY}^2} + \frac{\left(26.2/10^3 PY - 13.7/10^3 PY\right)^2}{108/10^6\ \mathrm{PY}^2} = 12.69$$

Test of Heterogeneity (IRD)
Test Statistic Interpretation

- Degrees of freedom for H = 2-1 = 1
- $Pr[\chi^2_1 > 12.69] = 0.0004$
- In a hypothesis testing framework with a pre-specified 2-sided alpha of 0.05, we reject the null hypothesis of no effect measure modification on the additive scale in these data. This interpretation hinges on the validity of our usual assumption regarding freedom from bias.
- These data are not very consistent with the state of nature described by the null. If the data arose from a single common rate difference, we would only expect to observe this degree of heterogeneity or more once in 25,000 such studies. For this interpretation to be valid, we need to assume no sources of bias.
- This is consistent with our less formal appraisal of the data. The rate difference is nearly tenfold greater in the older patients than in the younger patients, and there is relatively little overlap in the 95% confidence intervals.

BREAK

Week 2: Stratified Analysis & Effect Measure Modification

Video 10: Limitations and Reporting of Effect Measure Modification

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

Limitations of Tests of Homogeneity

- Tests of homogeneity share the same limitations as other hypothesis tests discussed earlier:
  - they fail to summarize the data with respect to their consistency with any alternative hypotheses
  - nor do they give us any indication of the *power* of the data to detect any alternative hypotheses of interest
- Considerably more data than are often available are needed to detect and characterize effect measure modification, if present.
- Many studies, unless explicitly designed to detect and characterize effect measure modification, will not have sufficient data to do so. Keep in mind that relatively precise stratum-specific estimates are usually required to statistically detect effect measure modification.
- Thus, not rejecting the null hypothesis of no effect measure modification may often be explained by a low statistical power.

Reporting Results

- When effect measure modification is present, it is not generally useful to summarize over strata with heterogeneous associations.

- Rather, it is of greater interest to report the observed associations by level of the modifier.

- If a summary measure is needed, weights that do not reflect arbitrary features of the study design should be chosen. Instead, ***standardization techniques*** (including ***inverse probability weighting***) should be used, with appropriate population-based weights (more on this later).

BREAK

Week 2: Stratified Analysis & Effect Measure Modification

Video 11: Effect Measure Modification on the Difference Scale with Multiplicative Models

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

Key Concepts

- Definition of effect measure modification
- Impact on generalizability (transportability)
- Scale dependence of effect measure modification
- Tests of homogeneity
  - Incidence rate ratio and difference
  - Cumulative incidence ratio and difference
  - Odds ratio
- Relative Excess Risk due to Interaction (RERI)

Can EMM on the Difference Scale be Detected when Estimation is Based on a Multiplicative Model?

- Epidemiologists often estimate associations on the multiplicative scale
  - Multiplicative models are often statistically efficient and convenient
    - Logistic regression
    - Cox Proportional Hazards regression
    - Poisson regression

- Detecting the presence of EMM on the multiplicative scale is straightforward
  - Stratified results
  - Incorporating multiplicative interaction terms in statistical models

- Whether EMM is also present on the additive scale may be of interest for several reasons
  - As an indication of mutual mechanistic interaction under the Rothman sufficient and component cause model
  - To develop intuition about absolute impact of exposure in subsets of the population

Asbestos, Smoking and Lung Cancer Incidence

10-year Cumulative Incidence of Lung Cancer
by Smoking and Asbestos Status in Telemark, Norway

	No Asbestos	Asbestos
Non-smoker	0.11%	0.67%
Smoker	0.95%	4.50%

- The CIR for the association between asbestos and lung cancer
  - Among non-smokers = 0.0067/0.0011 = 6.09
  - Among smokers =0.045/0.0095 = 4.74
- The CIR for the association between smoking and lung cancer
  - Among non-asbestos exposed = 0.0095/0.0011 = 8.64
  - Among asbestos exposed = 0.0450/0.0067 = 6.72
- From these data
  - Is asbestos more harmful among smokers or non-smokers?
  - Is smoking more harmful among those exposed or non-exposed to asbestos?

Asbestos, Smoking and Lung Cancer Incidence

10-year Cumulative Incidence of Lung Cancer
by Smoking and Asbestos Status in Telemark, Norway

	No Asbestos	Asbestos
Non-smoker	0.11%	0.67%
Smoker	0.95%	4.50%

- The CID for the association between asbestos and lung cancer
  - Among non-smokers = 0. 67% - 0. 11% = 0. 56%
  - Among smokers = 4.50% - 0. 95% = 3.55%
- The CID for the association between smoking and lung cancer
  - Among non-asbestos exposed = 0. 95% - 0. 11% = 0.84%
  - Among asbestos exposed = 4.50% - 0. 67% = 3.83%
- From these data
  - Is asbestos more harmful among smokers or non-smokers?
  - Is smoking more harmful among those exposed or non-exposed to asbestos?

Asbestos, Smoking and Lung Cancer Incidence

10-year Cumulative Incidence of Lung Cancer
by Smoking and Asbestos Status in Telemark, Norway

	No Asbestos	Asbestos
Non-smoker	0.11%	0.67%
Smoker	0.95%	??

Using the observed data, if there were no EMM on the additive scale, what would you expect to observe for the cumulative incidence among those exposed to both asbestos and smoking?

	No Asbestos	Asbestos
Non-smoker	0.11%	0.67%
Smoker	0.95%	4.5%

CID=0.56

Expect: 0.95%+0.56%
= 1.51%

CID=0.84

Expect: 0.67%+0.84%
= 1.51%

If there were no EMM on the additive scale, we would expect 1.51% among those exposed to asbestos and smoking. However, we observed 4.50%, indicating that the absolute effect of asbestos is stronger in the presence of smoking and vice-versa.

Can EMM on the Difference Scale be Detected when Estimation is Based on a Multiplicative Model?

- The Relative Excess Risk due to Interaction (RERI) is one metric that allows the assessment of EMM on the additive scale from multiplicative parameters

Relative Excess Risk due to Interaction (RERI)

Consider two exposures, A1 and A2:

	$A_1=0$	$A_1=1$
$A_2=0$	CI_{00}	CI_{10}
$A_2=1$	CI_{01}	CI_{11}

If there is no EMM on the additive scale then
$CI_{10} - CI_{00} = CI_{11} - CI_{01}$ and therefore, $CI_{11} - CI_{01} - CI_{10} + CI_{00} = 0$

Divide each term by CI_{00} and no EMM on the additive scale implies
$CIR_{11} - CIR_{01} - CIR_{10} + 1 = 0$.

This expression is called the Relative Excess Risk due to Interaction (RERI). The RERI allows the use of multiplicative parameters to determine whether there is EMM on the additive scale.

RERI = 0 if there is no EMM on the additive scale
RERI > 0 if the absolute effect of A1 is greater in the presence of A2 (and vice versa)
RERI < 0 if the absolute effect of A1 is weaker in the presence of A2 (and vice versa)

Relative Excess Risk due to Interaction (RERI)

10-year Cumulative Incidence Ratio of Lung Cancer
by Smoking and Asbestos Status in Telemark, Norway

	No Asbestos	Asbestos
Non-smoker	1	6.09
Smoker	8.64	40.91

$$RERI = CIR_{11}\text{-}CIR_{10}\text{-}CIR_{01}+1$$
$$RERI = 40.91 - 8.64 - 6.09 + 1$$
$$RERI = 27.18$$

- Using only results on the multiplicative scale
  - The RERI indicates that there is EMM on the additive scale
  - Asbestos is more harmful among smokers than non-smokers and
  - Smoking is more harmful among those exposed to asbestos than those not

BREAK