Suppose that we want to study the relationship between coffee drinking and heart attacks in adult males under 55. In particular, we want to know if there is an association between coffee drinking and heart attacks; and we also might like to estimate the odds of having a heart attack for the coffee drinkers versus the non coffee drinkers.

We could collect a sample of males under 55 and ask two questions: Have you had a heart attack ? Do you drink coffee ? If we constructed a 2 x 2 table with the data, what test would be appropriate to test for an association between coffee drinking and heart attack incidence ?

Suppose that we are concerned about the effects of smoking. How could we deal with this ?

To address this, we could obtain samples from smoking and non-smoking males under 55, conduct two 2 x 2 tests, and estimate two odds ratios. Suppose we did that:

For the smokers,

	Coffee		
	(Yes)	(No)	total
MI	1011	81	1092
No MI	390	77	467
total	1401	158	1559

For the non-smokers,

	Coffee		
	(Yes)	(No)	total
MI	383	66	449
No MI	365	123	488
total	748	189	937

Conduct the tests...

How should we combine the results ?

	Coffee		
	(Yes)	(No)	total
MI	1394	147	1541
No MI	755	200	955
total	2149	347	2496

Now what happens ? Conduct the test...

Simply combining the data defeats the purpose of controlling for the confounding variable ! The Mantel-Haenszel test can be used to assess the association of a dichotomous disease and a dichotomous exposure variable after controlling for k confounding variables (in our example, $k = 2$). To do this,

- Compute $O = \Sigma_{i=1}^{k} O_i$ in the (1,1) cell over the k strata.
- Compute $E = \Sigma_{i=1}^{k} E_i$ in the (1,1) cell over the k strata.
- Compute the variance of O: $V = \Sigma_{i=1}^{k} \frac{(a_i+b_i)(c_i+d_i)(a_i+c_i)(b_i+d_i)}{(n_i)^2(n_i-1)}$
- The test statistic is given by $\chi^2_{MH} = \frac{(|O-E|-0.5)^2}{V}$ which under H_0 is distributed as chi-square with 1 df.
- Note: This test should only be used when $V \geq 5$.

For our example, $O = 1394$, $E = 1339.76$ and $V = 67.5$. This gives $\chi^2_{MH} = 42.79$

As before, we might like to get estimates of the odds ratio:

For the smokers,

	Coffee (Yes)	Coffee (No)	total
MI	1011	81	1092
No MI	390	77	467
total	1401	158	1559

and $(O\hat{R}_S) = \frac{(1011)(77)}{(390)(81)} = 2.46$

For the non-smokers,

	Coffee (Yes)	Coffee (No)	total
MI	383	66	449
No MI	365	123	488
total	748	189	937

and $(O\hat{R}_{NS}) = \frac{(383)(123)}{(365)(66)} = 1.96$

Of course, what we really want to know is the true underlying OR. Mantel-Haenszel's Method can be used to combine data from multiple 2 x 2 tables to estimate the underlying odds ratio and test if that underlying odds ratio is significantly different than 1.

Before combining the data from the two tables, we should verify that the population odds ratios are not significantly different across strata. If they are, there is no good reason to try to obtain a common estimate.

We need to test the null $H_0 : OR_S = OR_{NS}$ against the alternative $H_A : OR_S \neq OR_{NS}$.

Recall that

$$\hat{se}(ln(O\hat{R})) = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$$

Let

$$\hat{se}^2(ln(O\hat{R})) = \frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}$$

$$\hat{se}_S^2 = \hat{se}^2(ln(O\hat{R}_S)) \text{ and } \hat{se}_{NS}^2 = \hat{se}^2(ln(O\hat{R}_{NS}))$$

Under the null,

$$X^2 = \frac{1}{\hat{se}_S^2}(ln(O\hat{R}_S) - ln(\hat{OR}))^2 + \frac{1}{\hat{se}_{NS}^2}(ln(O\hat{R}_{NS}) - ln(\hat{OR}))^2$$

is approximately chi-square with (2-1) degrees of freedom.

$$ln(\hat{OR}) = \frac{\frac{1}{\hat{se}_S^2} ln(O\hat{R}_S) + \frac{1}{\hat{se}_{NS}^2} ln(O\hat{R}_{NS})}{\frac{1}{\hat{se}_S^2} + \frac{1}{\hat{se}_{NS}^2}}$$

Recall the data

For the smokers,

	Coffee (Yes)	Coffee (No)	total
MI	1011	81	1092
No MI	390	77	467
total	1401	158	1559

and $(O\hat{R}_S) = \frac{(1011)(77)}{(390)(81)} = 2.46$

For the non-smokers,

	Coffee (Yes)	Coffee (No)	total
MI	383	66	449
No MI	365	123	488
total	748	189	937

and $(O\hat{R}_{NS}) = \frac{(383)(123)}{(365)(66)} = 1.96$

We need to test the null $H_0 : OR_S = OR_{NS}$ against the alternative $H_A : OR_S \neq OR_{NS}$.

$\hat{OR}_S = 2.46$ and $\hat{OR}_{NS} = 1.96$. So, $ln(\hat{OR}_S) = 0.9$ and $ln(\hat{OR}_{NS}) = 0.673$

$$\begin{aligned} \hat{se}^2_S &= \frac{1}{1011} + \frac{1}{390} + \frac{1}{81} + \frac{1}{77} \\ &= 0.02888504 \end{aligned}$$

and $\frac{1}{\hat{se}^2_S} = 34.62$.

Similarly,

$$\begin{aligned} \hat{se}^2_{NS} &= \frac{1}{383} + \frac{1}{365} + \frac{1}{66} + \frac{1}{123} \\ &= 0.02862869 \end{aligned}$$

and $\frac{1}{\hat{se}^2_{NS}} = 34.93$.

$$\begin{aligned} ln(\hat{OR}) &= \frac{\frac{1}{\hat{se}^2_S} ln(\hat{OR}_S) + \frac{1}{\hat{se}^2_{NS}} ln(\hat{OR}_{NS})}{\frac{1}{\hat{se}^2_S} + \frac{1}{\hat{se}^2_{NS}}} \\ &= \frac{34.62(0.9) + 34.93(0.673)}{34.62 + 34.93} \\ &= = 0.786 \end{aligned}$$

Under the null,

$$X^2 = \frac{1}{\hat{se}^2_S}(ln(\hat{OR}_S) - ln(\hat{OR}))^2 + \frac{1}{\hat{se}^2_{NS}}(ln(\hat{OR}_{NS}) - ln(\hat{OR}))^2$$

is approximately chi-square with (2-1) degrees of freedom.

Evaluating the test statistic gives

$$x^2 = (34.62)(0.9 - 0.786)^2 + (34.93)(0.673 - 0.786)^2 = 0.896$$

The null is not rejected (for $\alpha = 0.05$) and we can proceed. Recall that we want to combine the data from the two 2 x 2 tables to obtain a better estimate of the underlying odds ratio.

Using the Mantel-Haenszel method,

$$\hat{OR} = \frac{\frac{(ad)_S}{T_S} + \frac{(ad)_{NS}}{T_{NS}}}{\frac{(bc)_S}{T_S} + \frac{(bc)_{NS}}{T_{NS}}}$$

where T_S and T_{NS} represent the total number of observations in the tables obtained from smokers and non-smokers, repsectively.

For the data here,

$$\begin{aligned} \hat{OR} &= \frac{\frac{(1011)(77)}{1559} + \frac{(383)(123)}{937}}{\frac{(390)(81)}{1559} + \frac{(365)(66)}{937}} \\ &= 2.18 \end{aligned}$$

Conclusion: We estimate that males under the age of 55 who drink coffee have odds 2.18 times greater of having a heart attack than males who do not drink coffee.

An approximate 95 % confidence interval for the true odds ratio is given by

$$\left(e^{ln(\hat{OR})-1.96\hat{se}(ln(\hat{OR})),e^{ln(\hat{OR})-1.96\hat{se}(ln(\hat{OR}))}}\right)$$

This gives

$$\left(e^{0.786-1.96(0.120)}, e^{0.786+1.96(0.120)}\right) = (1.73, 2.78)$$