

## RESEARCH COMMUNICATION

# Assessment of Confounding and Interaction using the Mantel-Haenszel Risk Estimation Method

Sachin Jose, Preethi Sarah George, Aleyamma Mathew\*

### Abstract

The association between an exposure of interest (risk factor) and a disease may be confounded by the action of other separate factors as well as by interactions between risk factors exerting an impact. Crude measures of effect may be misleading in such situations. Levels of the potential confounding factor could be estimated using stratified analysis. Uniformity of the stratum-specific effect estimates can be assessed by performing chi-square tests for heterogeneity. If the effect is uniform across strata, we can calculate a pooled adjusted summary estimate of the effect using the Mantel-Haenszel (M-H) method. Confidence intervals for the adjusted estimate and the M-H chi-square test are calculated to assess the significance. If the effect is not uniform (presence of interaction), we report stratum-specific estimates, confidence intervals and chi-square for each estimate. In the present paper, assessment of the level of confounding and interaction between risk factors are illustrated using a case-control study of lung cancer conducted at the Regional Cancer Centre, Trivandrum.

**Key Words:** Confounding - interaction - Mantel-Haenszel method - lung cancer

*Asian Pacific J Cancer Prev*, 9, 323-326

### Introduction

In the analysis of exposure-outcome (disease of interest) association, we usually include all subjects in the study sample. This provides a crude estimate of the effect of the exposure on the outcome of interest. The association between the exposure of interest and the disease may be confounded by other factors (a confounder is an extraneous factor that satisfies the conditions such as (a) it is a risk factor for the disease & (b) it is associated with the study exposure but is not a consequence of exposure of different risk factors on the occurrence of the disease under study). Further, the disease may result from the interaction between risk factors (interaction or effect modification is concerned with the way in which two or more potential risk factors act together).

When we divide the study subjects into several subgroups or strata, defined by a potential confounding variable, and the stratum-specific results are not homogenous, we calculate a summary effect estimate in contrast to the crude estimate. Mantel-Haenszel (M-H) risk estimation procedure is used to estimate such summary estimates (Schlesselman 1974). The objectives of the present study were to assess the level of confounding and interaction between lung cancer risk factors using the M-H method. The illustration is based on an unmatched lung cancer case-control study and the factors considered for illustration are age, tobacco smoking and alcohol drinking. Tobacco smoking is thought to account for between 75%

and 90% of all lung cancers (Peto et al., 2006) and hard liquor consumption moderately increases the risk of lung cancer (De Stefani et al., 2002).

### Materials and Methods

The case-control study was conducted at the Regional Cancer Centre (RCC), Trivandrum during 2006. Primary male lung cancer cases histologically/ radiologically diagnosed and reported at the RCC during the above period were considered as 'cases'. Non-tobacco related male cancer patients (bone, prostate and thyroid cancers) reported at the same hospital during the same period were considered as 'controls'. In the present illustration, tobacco consumption was considered as risk factor of interest, age and alcohol consumption were considered as extraneous factors. A pre-tested self-administered questionnaire was used to collect information from each case and control after obtaining oral consent from each subject. The questionnaire-included information regarding age, tobacco consumption (smoking and chewing) and alcohol consumption of each case and control. Questionnaire was computerized and information was entered in the computer. Assuming tobacco prevalence =30% (Reddy and Gupta, 2004), odds ratio for tobacco smoking =3.0; alcohol prevalence =15%, odds ratio for alcohol drinking =2.5, power of the study=90%, type 1 error = 5%, a sample size of 150 cases and 150 controls are needed (Schlesselman, 1974).

**Results**

A total of 167 cases and 209 controls were obtained and all these subjects are included in the study. Men with lung cancer had a much higher prevalence of smoking (167/209= 80%) compared with the controls (45/169=27%) with an odds ratio (OR) of 11.0 (Table 1).

It is possible that the observed OR of 11.0 for smoking arose due to the confounding effect of other variables such as age, alcohol consumption etc. To assess whether age is a confounder, we looked at the association between smoking and lung cancer separately for age (Table 2). Among the controls, proportion of smokers with age more than or equal to 50 years were much higher (32/45=71.1%) than non-smokers (70/124=56.5) in the same age group, that is there was an association between smoking and age (Table 3).

The association between smoking and age was examined among the controls (Table 3). This is because controls should represent the population from which the cases were drawn and we need to assess that association

**Table 1. Association between Smoking and Lung Cancer (crude estimate)**

	Smoking		Total
	Yes	No	
Lung cancer cases	167 (a)	42 (b)	209 (n <sub>1</sub> )
Controls	45 (c)	124 (d)	169 (n <sub>0</sub> )
Total	212 (m <sub>1</sub> )	166 (m <sub>0</sub> )	378 (N)

Crude odds ratio (OR)  $ad/bc = 11.0$   
 Approximate 95 % CI  $\text{Exp} [\ln (\text{OR}) \pm 1.96 \text{ SE} (\ln \text{OR})]$   
 (OR) = (6.78, 17.71) Where  $\text{SE} (\ln \text{OR}) = \sqrt{(1/a + 1/b + 1/c + 1/d)}$   
 $\chi^2$  on 1 df = 107.39 (O-E)<sup>2</sup>/V where O= a, E =  $m_1 n_1 / N$   
 = 117.22 and  $V = [n_1 n_0 m_1 m_0] / [N^2 (N-1)] = 23.076$   
 P= 0.0000001

**Table 2. Association between Smoking and Lung Cancer Stratified by Age**

Age < 50 years	Smoking		Total
	Yes	No	
Lung cancer cases	17 (a <sub>1</sub> )	6 (b <sub>1</sub> )	23 (n <sub>11</sub> )
Controls	13 (c <sub>1</sub> )	54 (d <sub>1</sub> )	67 (n <sub>01</sub> )
Total	30 (m <sub>11</sub> )	60 (m <sub>01</sub> )	90 (N1)
Crude OR= 11.8	App 95 % CI (OR) = (3.81, 35.52)		

  

Age >= 50 years	Smoking		Total
	Yes	No	
Lung cancer cases	150 (a <sub>2</sub> )	36 (b <sub>2</sub> )	186 (n <sub>12</sub> )
Controls	32 (c <sub>2</sub> )	70 (d <sub>2</sub> )	102 (n <sub>02</sub> )
Total	182 (m <sub>12</sub> )	106 (m <sub>02</sub> )	288 (N2)
Crude OR=9.1	Approximate 95 % CI (OR) = (5.19, 15.64)		

**Table 3. Distribution of Controls by Smoking Habits and Age**

	Smoking	Age		Total	p-value
		< 50	≥ 50		
Smoking	Yes	13	32	45	0.07
	No	54	70		
Total		67	102	169	

in the general population. Subjects with age ≥ 50 years (186/209=89%) is considerably higher among lung cancer patients than among controls (102/169=60.3%) (Table 3). Since age in these data was associated with both the exposure (smoking) and the outcome of interest (lung cancer), it acted as a confounding factor.

Examination of the stratum-specific odds ratios (and their confidence intervals) suggested that the effect of smoking on lung cancer in men who had age < 50 years is similar to the effect in those who had age ≥ 50 years. This is confirmed by Chi-square test for heterogeneity (p=0.686). Since the effect of smoking on lung cancer is uniform across the two strata, it is appropriate to calculate a pooled adjusted odds ratio.

**Mantel-Haenszel Odds Ratio (ORM-H) Smoking adjusted for Age**

Since age is a confounder of the relationship between smoking and lung cancer, the crude odds ratio of 11.0 is not appropriate. We could calculate separate odds ratios and their 95% confidence intervals for each age group. To summarize the overall results of the study in a way that removes the confounding effect of age, Mantel-Haenszel (M-H) estimate would be more appropriate. The M-H odds ratio denoted  $OR_{MH}$ , gives a weighted average of the odds ratios in the different strata, where those from larger strata are given more weight.

Thus  $OR_{MH} = [a_1 d_1 / N_1 + a_2 d_2 / N_2] / [b_1 c_1 / N_1 + b_2 c_2 / N_2] = 9.6$  (Table 2)

**Confidence Interval (CI) & test of significance for  $OR_{MH}$**

95% CI for  $OR_{MH} = \text{Exp} [(\ln OR_{MH} \pm SE (\ln OR_{MH}))]$

where  $SE (\ln OR_{M-H}) = \sqrt{\frac{\sum (b_1 c_1 / N_1)^2 v_1}{\sum (b_1 c_1 / N_1)^2}}$   
 &  $v_1 = 1/a_1 + 1/b_1 + 1/c_1 + 1/d_1$

95% CI for  $OR_{MH} = (5.3, 17.3)$

M-H Chi-square on 1 df =  $[\_O(a_1) - \_E(a_1)]^2 / \_V (a_1) = 90.81$  & P-value = 0.000001

Thus the odds ratio for smoking adjusted (9.6) contrasts with the crude odds ratio of 11.0.

When the data of all ages were combined, the result suggested an association between smoking and lung cancer stronger than really existed (positive confounding). But even after adjusting for age, the 95% CI is still consistent with an effect of smoking on lung cancer. Thus these results suggest that smoking increased the risk of lung cancer but this risk was reduced after allowing for the confounding effect of age.

**Table 4. Distribution of Cases and Controls by Age**

	Age	Age		Total	p-value
		< 50	≥ 50		
Lung cancer cases	23	186	209	388	0.00001
Controls	67	102	169		
Total	90	288	378		

**Table 5. Smoking and Lung Cancer Stratified by Age Group and Alcohol**

Alcohol use	Smoking		Total
	Yes	No	
<b>No</b>			
Age<50 Years			
Lung Cancer Cases	9 (a <sub>1</sub> )	6 (b <sub>1</sub> )	15
Controls	6 (c <sub>1</sub> )	53 (d <sub>1</sub> )	59
Total	15	59	74
Crude OR=13.3	App 95% CI (OR)=(3.5,50.5)		
Age>=50 Years			
Lung Cancer Cases	79 (a <sub>2</sub> )	34 (b <sub>2</sub> )	113
Controls	21 (c <sub>2</sub> )	64 (d <sub>2</sub> )	85
Total	100	98	198
Crude OR=7.1	App 95% CI (OR)=(3.76,13.4)		
<b>Yes</b>			
Age<50 Years			
Lung Cancer Cases	8 (a <sub>3</sub> )	0*(b <sub>3</sub> )	8
Controls	7 (c <sub>3</sub> )	1 (d <sub>3</sub> )	8
Total	15	1	16
Crude OR=2.29	App 95% CI (OR)=(0.07,79.2)		
Age>=50 Years			
Lung Cancer Cases	71 (a <sub>4</sub> )	2 (b <sub>4</sub> )	73
Controls	11 (c <sub>4</sub> )	6 (d <sub>4</sub> )	17
Total	82	8	90
Crude OR=19.4	App 95% CI (OR)=(3.47,108.5)		

\*0.5 is used for estimating OR

*Using the Mantel-Haenszel method to adjust for several confounders*

The above same technique can be used to obtain an OR adjusted for more than one confounder. Alcohol consumption may have confounded the association between lung cancer and smoking. Thus we constructed four 2 x 2 tables of smoking and lung cancer, one for each age group (<50 vs. ≥ 50 years) and alcohol consumption (never users vs. users). The cells of these 2 x 2 tables are shown in Table 5.

*χ<sup>2</sup> test for heterogeneity*

$$\chi^2_w = \sum V_i^{-1} (\ln O\hat{R}_i - \ln O\hat{R}_w)^2 \sim \chi^2$$

with 3 d.f = 40.317 with p value =0.0000001

Examination of the stratum-specific odds ratios (and their CIs) suggested that the effect of smoking on lung cancer in men is different. This is confirmed by the Chi-square test for heterogeneity (p =0.000001). Since the effect of smoking on lung cancer is not uniform across the four strata, it is not appropriate to calculate a pooled adjusted OR.

**Discussion**

The magnitude of confounding can be evaluated by observing the degree of discrepancy between crude and adjusted estimates. A large difference indicates the presence of confounding and implies that the adjusted summary measure is a better estimate of the effect of the exposure on the outcome of interest than the crude summary measure, since it removes the effect of the

confounder. The M-H risk estimation procedure can be extended to more than two strata. It is possible to adjust for several confounders.

If there is substantial variation between the stratum-specific estimates of effect (interaction), M-H estimate of effect will be misleading, as it does not convey the full form of the exposure-outcome association, that is, that it varies according to the level of the stratifying variable. This method becomes impractical because most strata will have very sparse data when the number of confounders increases.

Another with the M-H method is that each explanatory variable included in the analysis has to be classified as either an exposure or a confounder, and there may be only one exposure. In the present analysis smoking was our exposure variable, and age and alcohol were confounders. We obtained an OR for smoking adjusted for age (adjusted for alcohol not calculated as chi-square test for heterogeneity was significantly different). The results did not give us the OR for age adjusted for smoking. This would have required further M-H analysis.

**References**

De Stefani E, Correa P, Deneo-Pellegrini H, et al (2002). Alcohol intake and risk of adenocarcinoma of the lung. A case-control study in Uruguay. *Lung Cancer*, **38**, 9-14.

Peto RR, Lopez AD, Boreham J (2006). Mortality from smoking in developed countries. 1950-2000: Indirect Estimates from National Vital Statistics.

Reddy KS, Gupta PC (2004). Report on Tobacco Control in India. Ministry of Health and Family Welfare, New Delhi, Government of India.

Schlesselman JJ (1974). Case-control Studies. Design, conduct and analysis, Oxford university press, Oxford.

