

Lecture 4: Confounding and Effect Modification

Dankmar Böhning

Southampton Statistical Sciences Research Institute
University of Southampton, UK

March 2 - 4, 2015

Occurrence of confounding

Dealing with confounding

Effect modification

extensions

disease, exposure and stratum

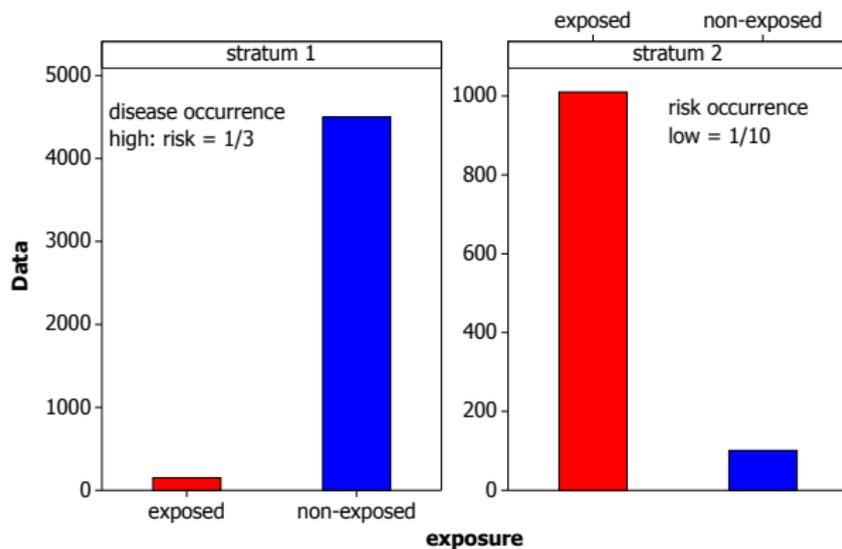
	exposed		non-exposed		
stratum	case	non-case	case	non-case	<i>RR</i>
1	50	100	1500	3000	1
2	10	1000	1	100	1
total	60	1100	1501	3100	0.1585

explanation?

explanation

	exposed		non-exposed		
stratum	case	non-case	case	non-case	<i>RR</i>
1	50	100	1500	3000	1
2	10	1000	1	100	1
total	60	1100	1	3	0.1585

	exposed	non-exposed	
stratum	risk	risk	<i>RR</i>
1	1/3	1/3	1
2	1/10	1/10	1
total	1/20	1/3	0.1585



conditions for confounding

- ▶ different amount of disease occurrence in the strata
- ▶ different exposure distributions in the strata

coffee consumption, smoking and coronary heart disease

in a cohort study 1,000 coffee drinkers and 1,000 non-drinkers were observed with respect to the development of coronary heart disease

	exposed (coffee)		non-exposed		
stratum	case	non-case	case	non-case	<i>RR</i>
smoking	195	705	21	79	1.03
non-smoking	5	95	29	871	1.55
total	200	800	50	950	4

coffee consumption, smoking and coronary heart disease

in a cohort study 1,000 coffee drinkers and 1,000 non-drinkers were observed with respect to the development of coronary heart disease

stratum	exposed (coffee)		non-exposed		RR
	case	non-case	case	non-case	
smoking	195	705	21	79	1.03
non-smoking	5	95	29	871	1.55
total	200	800	50	950	4

forms of confounding

- ▶ **inflation:** crude RR is larger (in an absolute sense) than the stratified RR
- ▶ **masking:** crude RR is smaller than stratified RR
- ▶ **effect modification:** crude RR is between stratified RR

dealing with confounding

stratify:

	exposed		non-exposed		
stratum	case	non-case	case	non-case	RR
1	$Y_1^{(1)}$	$n_1^{(1)} - Y_1^{(1)}$	$Y_0^{(1)}$	$n_0^{(1)} - Y_0^{(1)}$	$\widehat{RR}^{(1)}$
2	$Y_1^{(2)}$	$n_1^{(2)} - Y_1^{(2)}$	$Y_0^{(2)}$	$n_0^{(2)} - Y_0^{(2)}$	$\widehat{RR}^{(2)}$
...
k	$Y_1^{(k)}$	$n_1^{(k)} - Y_1^{(k)}$	$Y_0^{(k)}$	$n_0^{(k)} - Y_0^{(k)}$	$\widehat{RR}^{(k)}$

$$\widehat{RR}^{(j)} = \frac{Y_1^{(j)} / n_1^{(j)}}{Y_0^{(j)} / n_0^{(j)}}$$

dealing with confounding

evidently a solution could be a weighted estimator of the stratified relative risk estimators:

$$\widehat{RR} = \frac{\sum_j w_j \widehat{RR}^{(j)}}{\sum_j w_j}$$

but which weights $w_j > 0$ to choose?

the Mantel-Haenszel estimator

note that

$$\widehat{RR}^{(j)} = \frac{Y_1^{(j)} n_0^{(j)} / n^{(j)}}{Y_0^{(j)} n_1^{(j)} / n^{(j)}}$$

where $n^{(j)} = n_1^{(j)} + n_0^{(j)}$ is the size of the stratum

now consider the **Mantel-Haenszel estimator**:

$$\widehat{RR}_{MH} = \frac{\sum_j Y_1^{(j)} n_0^{(j)} / n^{(j)}}{\sum_j Y_0^{(j)} n_1^{(j)} / n^{(j)}}$$

advantages of the Mantel-Haenszel estimator

- ▶ principle of MH-estimation: sums before ratios
- ▶ MHE is not affected by zero-counts in the strata
- ▶ suitable for refined stratification and sparse data

Mantel-Haenszel estimator as weighted estimator

we have

$$\widehat{RR}^{(j)} = \frac{Y_1^{(j)} n_0^{(j)} / n^{(j)}}{Y_0^{(j)} n_1^{(j)} / n^{(j)}}$$

use the **Mantel-Haenszel weights** $w_j = Y_0^{(j)} n_1^{(j)} / n^{(j)}$ and see that

$$\begin{aligned} & \frac{\sum_j w_j \widehat{RR}^{(j)}}{\sum_j w_j} \\ &= \frac{\sum_j Y_1^{(j)} n_0^{(j)} / n^{(j)}}{\sum_j Y_0^{(j)} n_1^{(j)} / n^{(j)}} = \widehat{RR}_{MH} \end{aligned}$$

cs - Cohort studies

Main f/in Weights Options

Stratify on variables: S

Number of subjects variable:

Within-stratum weights

Use Mantel-Haenszel

Use external

Use internal

User-specified variable:

Report odds ratio

Woolf approximation

Test-based conf. intervals

Fisher's exact p

95 Confidence level

Include missing categories

Display pooled estimate

No crude estimate

No homogeneity test

Calculate standardized risk difference

OK Cancel Submit

```
(1 real change made)
. replace var10 = 79 in 4
(1 real change made)
. replace var10 = 5 in 5
(1 real change made)
. replace var10 = 95 in 6
(1 real change made)
. replace var10 = 29 in 7
(1 real change made)
. replace var10 = 871 in 8
(1 real change made)
. rename var10 freq
. cs D E [fweight = freq], by(S)
```

S	RR	[95% Conf. Interval]		M-H Weight
0	1.031746	.6916489	1.539076	18.9
1	1.551724	.6144943	3.918422	2.9
Crude	4	2.971453	5.384571	
M-H combined	1.100917	.7633712	1.587719	

Test of homogeneity (M-H) $\chi^2(1) = 0.629$ $\text{Pr}>\chi^2 = 0.4279$

Command

effect modification

- ▶ we need to distinguish **confounding** from **effect modification** !
- ▶ **confounding** occurs when the stratified relative risk estimators are homogeneous but different from the crude relative risk estimator
- ▶ **effect modification** occurs when the stratified relative risk estimators are not homogeneous (the MH-estimator may or may not be similar to the crude estimator)

diagnosis of effect modification

hypothesis of homogeneity:

$$H_0 : RR^{(1)} = \dots = RR^{(k)}$$

$H_1 : H_0$ is wrong

test statistics

$$\chi_{k-1}^2 = \sum_{j=1}^k \frac{(\log \widehat{RR}^{(j)} - \log \widehat{RR}_{MH})^2}{\text{Var}(\log \widehat{RR}^{(j)})}$$

if this test statistic is significant we reject H_0

cs - Cohort studies

Main f/in Weights Options

Stratify on variables: S

Number of subjects variable:

Within-stratum weights

Use Mantel-Haenszel

Use external

Use internal

User-specified variable:

Include missing categories

Display pooled estimate

No crude estimate

No homogeneity test

Calculate standardized risk difference

Report odds ratio

Woolf approximation

Test-based conf. intervals

Fisher's exact p

95 Confidence level

OK Cancel Submit

```
(1 real change made)
. replace var10 = 79 in 4
(1 real change made)
. replace var10 = 5 in 5
(1 real change made)
. replace var10 = 95 in 6
(1 real change made)
. replace var10 = 29 in 7
(1 real change made)
. replace var10 = 871 in 8
(1 real change made)
. rename var10 freq
. cs D E [fweight = freq], by(S)
```

	S	RR	[95% Conf. Interval]		M-H Weight
	0	1.031746	.6916489	1.539076	18.9
	1	1.551724	.6144943	3.918422	2.9
Crude		4	2.971453	5.384571	
M-H combined		1.100917	.7633712	1.587719	

Test of homogeneity (M-H) chi2 (1) = 0.629 Pr>chi2 = 0.4279

Command

coffee consumption, smoking and coronary heart disease

in a cohort study 1,000 coffee drinkers and 1,000 non-drinkers were observed with respect to the development of coronary heart disease

stratum	exposed (coffee)		non-exposed		χ^2
	case	non-case	case	non-case	
smoking	195	705	21	79	0.101
non-smoking	5	95	29	871	0.527
total	200	800	50	950	0.629

Mantel-Haenszel estimation for other effect measures

Mantel-Haenszel estimation can easily be extended to other effect measure such as if we have person-time data for stratum j :

$$\widehat{RR}^{(j)} = \frac{Y_1^{(j)} / T_1^{(j)}}{Y_0^{(j)} / T_0^{(j)}}$$

Mantel-Haenszel estimator:

$$\widehat{RR}_{MH} = \frac{\sum_j Y_1^{(j)} T_0^{(j)} / T^{(j)}}{\sum_j Y_0^{(j)} T_1^{(j)} / T^{(j)}}$$

where $T^{(j)} = T_1^{(j)} + T_0^{(j)}$ is the total person-time of the stratum j

Mantel-Haenszel estimation for other effect measures

Mantel-Haenszel estimation for odds ratio

in stratum j :

$$\widehat{OR}^{(j)} = \frac{Y_1^{(j)} / (n_1^{(j)} - Y_1^{(j)})}{Y_0^{(j)} / (n_0^{(j)} - Y_0^{(j)})}$$

Mantel-Haenszel estimator:

$$\widehat{OR}_{MH} = \frac{\sum_j Y_1^{(j)} (n_0^{(j)} - Y_0^{(j)}) / n^{(j)}}{\sum_j Y_0^{(j)} (n_1^{(j)} - Y_1^{(j)}) / n^{(j)}}$$

where $n^{(j)} = n_1^{(j)} + n_0^{(j)}$ is the size the stratum j

case-control study on lip cancer and sun exposure

a case-control study investigated the effect of sun exposure on lip cancer treating smoking as a confounder

	smoking		non smoking		
disease status	sun	no sun	sun	no sun	size
case	51	24	15	3	29
control	6	10	8	5	93

Lecture 4: Confounding and Effect Modification

extensions

The screenshot shows the Stata/IC 12.1 interface. The main window displays the results of a case-control study analysis. The dialog box 'cc - Case-control studies' is open, showing options for stratification and confidence intervals.

cc - Case-control studies

Main | **f/in** | Weights | Options

Stratify on variable: [dropdown] Number of subjects variable: [dropdown]

Within stratum weights

Use Mantel-Haenszel Exact confidence intervals

Use external Cornfield approximation

Use internal Woolf approximation

User-specified variable: [dropdown] Test-based confidence intervals

Include missing categories Fisher's exact p

Display pooled estimate Confidence level: [dropdown]

No crude estimate

No homogeneity test

Breslow-Day homogeneity test

Tarone's homogeneity test

OK Cancel Submit

Command

```

55 replace var10 = ...
56 replace var10 = ...
57 replace var10 = ...
58 rename var10 f...
59 cs D E [fweight...
60 cs D E [fweight...
61 generate var11 = ...
62 replace var11 = ...
63 replace var11 = ...
64 replace var11 = ...
65 replace var11 = ...
66 replace var11 = ...
67 replace var11 = ...
68 replace var11 = ...
69 rename var11 f...
70 cc D E [fweight...
71 cc D E [fweight...

```

	S	OR	[95% Conf. Interval]	M-H Weight
	0	3.541667	1.011455 13.14962	1.582418 (exact)
	1	3.125	.4483337 24.66091	.7741935 (exact)
Crude		2.619048	1.016247 6.717228	(exact)
M-H combined		3.404783	1.341535 8.641258	

Test of homogeneity (M-H) chi2(1) = 0.01 Pr>chi2 = 0.9029

Test that combined OR = 1:

Mantel-Haenszel chi2(1) = 6.96
Pr>chi2 = 0.0083

Variables

Variable	Label
Y	
P_time	
x	
logP_time	
var6	
D	
E	
S	
freq	
freq_case_co...	

Properties

Variables

Name	P_time

Command