

# Practical 2: Confounding and Effect Modification for Effect Measures in Cohort and Case-Control Studies

## *Solutions*

Dankmar Böhning  
 Southampton Statistical Sciences Research Institute  
 University of Southampton, UK

Advanced Statistical Methods in Epidemiology

### 1 Cohort Studies: The BELCAP Study

**questions:** Calculate the

- **crude** relative risk
- and **Mantel-Haenszel** relative risk (adjusting for initial dental status)

event is IMPROVEMENT (DMFT at begin of study **larger** than DMFT at the end of study), so that we are considering relative risk for improvement of dental status with and without adjusting for dental status at begin of study

use data file `improve.dat`; results are relative risk evaluation for school 1, school 2, school 4, school 5, and school 6

```
. cs Improve1 school1, by(dmftb)
```

dmftb	RR	[95% Conf. Interval]		M-H Weight
0	.	.	.	0
1	.7222222	.0933458	5.587877	.9473684
2	1.54386	.7367993	3.234941	3.257143
3	.7384615	.4226078	1.290382	5
4	1.209091	.94215	1.551665	8.04878
5	1.25	1.004021	1.556242	6.588235
6	1.076923	.9312996	1.245317	6.933333
7	1	1	1	5.454545
8	1	1	1	4.736842

Crude	1.147787	.9705709	1.357361
M-H combined	1.099201	.9761553	1.237757

-----  
 Test of homogeneity (M-H)      chi2(7) =      4.882   Pr>chi2 = 0.6743

cs Improve1 school2, by(dmftb)

dmftb	RR	[95% Conf. Interval]		M-H Weight
0	.	.	.	0
1	2.708333	.8763556	8.36997	1.142857
2	1.62963	.7836907	3.3887	3.176471
3	.8461538	.5640079	1.269444	6.5
4	.9212121	.599275	1.416098	5.5
5	1.25	1.004021	1.556242	4.571429
6	1.076923	.9312996	1.245317	6.259259
7	1	1	1	4.117647
8	1	1	1	3.75

Crude	.8591833	.7002344	1.054213
M-H combined	1.118323	.9731964	1.285091

-----  
 Test of homogeneity (M-H)      chi2(7) =      7.224   Pr>chi2 = 0.4059

. cs Improve1 school4, by(dmftb)

dmftb	RR	[95% Conf. Interval]		M-H Weight
0	.	.	.	0
1	1.083333	.2935185	3.998422	1.655172
2	.2962963	.0420228	2.089138	2.16
3	.967033	.674078	1.387306	6.066667
4	1.097778	.8088136	1.48998	6.617647
5	1.071429	.7888389	1.455252	6.588235
6	1.076923	.9312996	1.245317	6
7	1	1	1	5
8	1	1	1	5.238095

Crude	.8865398	.7272396	1.080734
M-H combined	.9999267	.8813615	1.134442

-----  
 Test of homogeneity (M-H)      chi2(7) =      3.093   Pr>chi2 = 0.8763

cs Improve1 school5, by(dmftb)

dmftb	RR	[95% Conf. Interval]		M-H Weight
0	.	.	.	0
1	2.294118	.7721171	6.816293	1.7
2	1.454545	.6333869	3.3403	2.444444
3	.8951049	.5812851	1.378347	5.296296
4	1.013333	.7187958	1.428562	6.617647
5	1.193182	.9410845	1.512811	8.380952
6	1.076923	.9312996	1.245317	7.129032
7	1	1	1	5.652174
8	1	1	1	5

Crude	.9794449	.8198088	1.170166
M-H combined	1.118692	.9902527	1.263789

Test of homogeneity (M-H)       $\chi^2(7) = 3.944$      $\text{Pr} > \chi^2 = 0.7862$

cs Improve1 school6, by(dmftb)

dmftb	RR	[95% Conf. Interval]		M-H Weight
0	.	.	.	0
1	2.333333	.7668385	7.099858	1.5
2	1.73913	.8641024	3.50025	3.538462
3	.852071	.5530998	1.312647	5.827586
4	.974359	.6679431	1.421342	6.09375
5	.75	.3548161	1.585328	3.2
6	.9940828	.8026919	1.231108	6.259259
7	1	1	1	5.652174
8	1	1	1	1.666667

Crude	.912838	.7486913	1.112973
M-H combined	1.081808	.9166379	1.27674

Test of homogeneity (M-H)       $\chi^2(7) = 6.591$      $\text{Pr} > \chi^2 = 0.4727$

## 2 Cohort study on risk of developing a respiratory disease

Swan (1986) gives the following data from a study of infant respiratory disease. The numbers show the proportion of children developing bronchitis or pneumonia in their first year of life by sex and type of feeding.

Table 1: Numbers are: number with bronchitis or pneumonia / at risk

Sex	type of feeding		
	Bottle only	Breast +supplement	Breast only
Boys	77/458	19/147	47/494
Girls	48/384	16/127	31/464

- what could be the risk/preventive factor here?
- what could be a confounder?

data are organized in this way; since there are two exposure categories two analyses need to be calculated!

```

+-----+
Case  Exposure    Sex  freq_B_S  freq_b~y
-----+-----
1. 1      0    boys      47        47
2. 0      0    boys     447       447
3. 1      1    boys      19        77
4. 0      1    boys     128       381
5. 1      0   girls      31        31
-----+-----
6. 0      0   girls     433       433
7. 1      1   girls      16        48
8. 0      1   girls     111       336
+-----+

```

for breast and supplement:

```
. cs Case Exposure [fweight = freq_B_S], by(Sex)
```

Sex	RR	[95% Conf. Interval]	M-H Weight
boys	1.358518	.8239773 2.239832	10.77847
girls	1.8857	1.065867 3.336124	6.661591
Crude	1.568875	1.077767 2.283767	
M-H combined	1.559886	1.072168 2.269462	

```
Test of homogeneity (M-H)      chi2(1) = 0.718 Pr>chi2 = 0.3967
```

for bottle only:

Table 2: numbers are: survivors/non-survivors

Sex	class traveled			
	First	Second	Third	Crew
M	57/118	14/154	75/387	192/670
W/C	146/4	104/13	103/141	20/3

```
. cs Case Exposure [fweight = freq_bottle_only], by(Sex)
```

Sex	RR	[95% Conf. Interval]		M-H Weight
boys	1.767072	1.258011	2.482128	22.61134
girls	1.870968	1.216	2.878718	14.03774
Crude	1.823345	1.395773	2.381897	
M-H combined	1.806868	1.383733	2.359392	

Test of homogeneity (M-H)      chi2(1) =      0.042    Pr>chi2 = 0.8383

### 3 Survival Analysis on the S.S. Titanic

At 2.20 a.m. on the morning of April 15th 1912 the unsinkable S.S. Titanic sank in the North Atlantic Ocean with the loss of nearly 1500 lives. The Titanic represented the state of the art in the shipping technology of its day. It was 852 ft long, displaced 52,310 tons, and was equipped with a system of sealed bulkheads which were believed to render the vessel unsinkable. While on its maiden voyage the Titanic struck an iceberg producing a 300 ft long gash in its side and flooding five bulkheads. In the three hours after the ship struck the iceberg, the Titanics bulkheads gradually filled with water and it sank before a rescue ship, the S.S. Carpathia, could reach it. Two-thirds of the Titanics complement of passengers and crew went to the bottom with it.

#### 3.1 Survival Analysis on the S.S. Titanic

- provide evidence that the risk for surviving was considerably lower for men than for woman and children (W/C)!
- provide an analysis for the survival risk of the people based on their class traveled! Use gender as a potential confounder!

Let us list the data at first:

```
. list survival SocialClass Gender frequency, abbreviate(20)
```

	survival	SocialClass	Gender	frequency
1.	survived	first	man	57
2.	died	first	man	118
3.	survived	second	man	14
4.	died	second	man	154
5.	survived	third	man	75
6.	died	third	man	387
7.	survived	crew	man	192
8.	died	crew	man	670
9.	survived	first	woman/child	146
10.	died	first	woman/child	4
11.	survived	second	woman/child	104
12.	died	second	woman/child	13
13.	survived	third	woman/child	103
14.	died	third	woman/child	141
15.	survived	crew	woman/child	20
16.	died	crew	woman/child	3

. list survival SocialClass Gender freq\_second freq\_third freq\_crew, abbreviate(20)

	survival	SocialClass	Gender	freq_second	freq_third	freq_crew
1.	survived	first	man	57	57	57
2.	died	first	man	118	118	118
3.	survived	second	man	14	.	.
4.	died	second	man	154	.	.
5.	survived	third	man	.	75	.
6.	died	third	man	.	387	.
7.	survived	crew	man	.	.	192
8.	died	crew	man	.	.	670
9.	survived	first	woman/child	146	146	.
10.	died	first	woman/child	4	4	.
11.	survived	second	woman/child	104	.	.
12.	died	second	woman/child	13	.	.
13.	survived	third	woman/child	.	103	.
14.	died	third	woman/child	.	141	.
15.	survived	crew	woman/child	.	.	20

16.	died	crew	woman/child	.	.	3
-----	------	------	-------------	---	---	---

first we look at survival risk for men:

```
. cs survival Gender [fweight = frequency], by(SocialClass)
```

SocialClass	RR	[95% Conf. Interval]		M-H Weight
first	.334638	.26995	.4148271	78.61538
second	.09375	.0565449	.1554351	61.30526
third	.3845669	.2983403	.4957148	67.40227
crew	.2561485	.2094013	.3133316	19.48023
Crude	.2902776	.2599628	.3241275	
M-H combined	.2776221	.2400445	.3210823	

Test of homogeneity (M-H)      chi2(3) = 27.557   Pr>chi2 = 0.0000

Clearly, almost 4 times lower that for women and children!

Then, we look at class traveled (using first class as reference):

```
. cs survival SocialClass [fweight = freq_second], by(Gender)
```

Gender	RR	[95% Conf. Interval]		M-H Weight
woman/child	.913242	.8520774	.9787972	63.97753
man	.255848	.1483592	.441214	27.91837
Crude	.6628641	.5638392	.7792802	
M-H combined	.7135229	.6371696	.7990257	

Test of homogeneity (M-H)      chi2(1) = 62.290   Pr>chi2 = 0.0000

```
cs survival SocialClass [fweight = freq_third], by(Gender)
```

Gender	RR	[95% Conf. Interval]		M-H Weight
woman/child	.4336964	.3735934	.5034686	90.41624
man	.4984051	.3702507	.6709175	41.34066

```

          Crude | .4036478 .3465698 .4701263
M-H combined | .4539997 .3947457 .5221481
-----
Test of homogeneity (M-H)      chi2(1) = 0.740 Pr>chi2 = 0.3897

. cs survival SocialClass [fweight = freq_crew], by(Gender)

          Gender |          RR      [95% Conf. Interval]      M-H Weight
-----+-----
      woman/child |          .          .          .          0
          man | .6838442 .5341948 .8754164 47.38091
-----+-----
          Crude | .7354545 .5765905 .9380891
M-H combined | .6838442 .5341948 .8754164
-----

```

### 3.2 Survival Analysis on the S.S. Titanic: the role of English language

An inquiry investigating the sinking of the ship was held by Lord Mersey in 1912, also called the Mersey inquiry. It was suggested in this investigation that the low survival rate of the third class was due to the fact that a large proportion of the third class travelers were emigrants and 'that their lack of English prevented them from following the crew's instructions'. The following table carries information on survival rates by nationality. Use these data to provide evidence that the above claim cannot be sustained. This analysis is more difficult since, for example, a simple comparison of British and non-British travelers – even if adjusted for gender – will not suffice since, if these risks are comparable one might argue that a selection process took place in the sense that those with good knowledge of English language had higher chances of survival whereas, if the risk for survival for British nationals turned out to be higher, this could be taken as reasons for the validity of the above claim. Hence a comparison with the Irish nationals is helpful who are native English speakers and many of them traveled in third class.

```

. cs survive nationality [fweight = freq_non_british], by(sex)

          sex |          RR      [95% Conf. Interval]      M-H Weight
-----+-----
      woman/child | .9780564 .6660151 1.436295 15.19048
          man | 1.541176 .9367743 2.535536 11.83295
-----+-----
          Crude | 1.218195 .8826072 1.681382

```

Table 3: numbers are: survivors/non-survivors

Sex	nationality		
	British	Non-British	Irish
M	17/114	60/240	8/41
W/C	22/30	48/68	32/32

```

M-H combined |      1.224634      .8968035      1.672304
-----
Test of homogeneity (M-H)      chi2(1) =      2.134 Pr>chi2 = 0.1440

. cs survive nationality [fweight = freq_irish], by(sex)

      sex |      RR      [95% Conf. Interval]      M-H Weight
-----+-----
  woman/child |      1.181818      .7914447      1.76474      12.13793
      man |      1.258103      .5804353      2.72696      4.627778
-----+-----
      Crude |      1.660994      1.143226      2.413259
M-H combined |      1.202875      .8382013      1.726206
-----
Test of homogeneity (M-H)      chi2(1) =      0.020 Pr>chi2 = 0.8865

```

Clearly, the survival risk for non-British is similar the Irish (if adjusted for gender).

## 4 Case-Control Studies

### 4.1 Source Investigations

During January 1984 six cases of legionnaires' disease were reported to the health authority in Reading, UK, all of whom became ill between 15 and 19 December 1983 (Anderson *et al.* 1985). This cluster suggested a point source outbreak. A local search was then conducted to discover whether there had been any other legionnaires' disease with onsets within the same 5 days. General practitioners and hospital physicians were asked to consider recent referrals and report any possible cases for further consideration, whilst hospital discharge and autopsy records were reviewed and outpatient X-rays were checked for possible cases, so far un-diagnosed.

The result was that 13 cases were detected in all. The cases had no obvious factor in common, such as all working in the same place, so that no clear source of the legionella bacterium was apparent. However, all cases had visited

Table 4: *Number of people visiting parts of Reading town center in the 2 weeks preceding the onset of legionnaires' disease*

Aread of Reading	Cases	Controls
Abbey Square	9	19
Butts Centre	12	21
Forbury Gardens	3	6
Minster Street	9	21
South Street	4	9
Railway Station	3	9
Overall	13	36

Reading town centre just before their illness.

A case-control study was mounted to compare exposure between the cases and a selected set of 36 people without the disease (the controls). Cases and controls were compared by the number who had visited each of six designated parts of Reading town centre, just prior to the outbreak. Results are given in the Table below.

In this investigation the case-control approach is the only one possible. Speed of hypothesis testing was of paramount importance since further infections could lead to deaths. Cohort studies of legionnaires' disease will, in any cases, be impractical because of the rare nature of the disease (only 558 cases were reported in England and Wales in the four years prior to the Reading outbreak), despite the common presence of the bacterium in many natural and human-made water supplies. Hence extremely large samples would be required.

**Task:** Identify the spatial source of contamination by using the tool of a case-control study and by computing the ORs with 95% CI!

```
. cc case_con expos [fweight = AbbeySquare]
```

	Exposed	Unexposed	Total	Proportion Exposed
Cases	9	4	13	0.6923
Controls	19	17	36	0.5278
Total	28	21	49	0.5714
	Point estimate		[95% Conf. Interval]	
Odds ratio	2.013158		.4485414	10.48914 (exact)
Attr. frac. ex.	.503268		-1.229449	.9046633 (exact)
Attr. frac. pop	.3484163			

chi2(1) = 1.06 Pr>chi2 = 0.3042

. cc case\_con expos [fweight = AbbeySquare ButtsCentre]

. cc case\_con expos [fweight = ButtsCentre]

	Exposed	Unexposed	Total	Proportion Exposed
Cases	12	1	13	0.9231
Controls	21	15	36	0.5833
Total	33	16	49	0.6735
	Point estimate		[95% Conf. Interval]	
Odds ratio	8.571429		1.014893	390.3122 (exact)
Attr. frac. ex.	.8833333		.0146748	.9974379 (exact)
Attr. frac. pop	.8153846			

chi2(1) = 5.01 Pr>chi2 = 0.0252

. cc case\_con expos [fweight = ForburyGardens]

	Exposed	Unexposed	Total	Proportion Exposed
Cases	3	10	13	0.2308
Controls	6	30	36	0.1667
Total	9	40	49	0.1837
	Point estimate		[95% Conf. Interval]	
Odds ratio	1.5		.2030941	8.687827 (exact)
Attr. frac. ex.	.3333333		-3.923827	.8848964 (exact)
Attr. frac. pop	.0769231			

chi2(1) = 0.26 Pr>chi2 = 0.6089

. cc case\_con expos [fweight = SouthStreet]

	Exposed	Unexposed	Total	Proportion Exposed
Cases	9	4	13	0.6923
Controls	21	15	36	0.5833

```

Total |          30          19 |          49          0.6122
      |          Point estimate |          [95% Conf. Interval]
-----+-----
Odds ratio |          1.607143 |          .3568467  8.442998 (exact)
Attr. frac. ex. |          .3777778 |          -1.802323  .8815587 (exact)
Attr. frac. pop |          .2615385 |
-----+-----
                                chi2(1) =          0.48 Pr>chi2 = 0.4894

```

```
. cc case_con expos [fweight = RailwayStation]
```

```

                                Proportion
                                Exposed
-----+-----
Cases |          3          10 |          13          0.2308
Controls |          9          27 |          36          0.2500
-----+-----
Total |          12          37 |          49          0.2449
      |          Point estimate |          [95% Conf. Interval]
-----+-----
Odds ratio |          .9 |          .1307846  4.687493 (exact)
Prev. frac. ex. |          .1 |          -3.687493  .8692154 (exact)
Prev. frac. pop |          .025 |
-----+-----
                                chi2(1) =          0.02 Pr>chi2 = 0.8901

```

The analysis clearly identifies Butts Centre as the source of contamination with significant OR=8.57.

## 5 Riskfactor for caries

In a case-control of risk factors for dental caries McMahon *et al.* (1993) present data relating to age of child, age of mother and whether or not the child has dental caries, as shown below. Caries is defined here as a minimum of four decayed, missing or filled teeth.

**Task:** Calculate the *crude* and *Mantel-Haenszel* odds ratio for caries of the child using a suitable reference group! Which age-variable is considered to be the confounder here?

the younger mothers:

```
. cc Case Exposure [fweight = fyoung], by(Age_Cild)
```

Table 5: *Frequencies by Age-of-Mother, Age-of-Child and Case-Control Status*

Age of mother (in yrs)	Age of child (months)					
	< 36		36-47		≥ 48	
	Caries	Control	Car.	Contr.	Car.	Contr.
< 25	1	1	1	5	8	9
25-34	4	16	18	67	46	113
≥ 35	1	2	3	14	10	35

Age_Cild	OR	[95% Conf. Interval]	M-H Weight
1	2	.0127932 234.4954	.2 (exact)
2	.9333333	.0148468 15.32852	.6521739 (exact)
3	3.111111	.799703 11.8356	1.451613 (exact)
Crude	2.428571	.7856042 7.291572	(exact)
M-H combined	2.398149	.8785518 6.546136	

Test of homogeneity (M-H)      chi2(2) =      0.75    Pr>chi2 = 0.6872

Test that combined OR = 1:

Mantel-Haenszel chi2(1) =      2.87

Pr>chi2 =      0.0900

the middle-aged mothers:

. cc Case Exposure [fweight = fmid], by(Age\_Cild)

Age_Cild	OR	[95% Conf. Interval]	M-H Weight
1	.5	.0212917 36.79341	.6956522 (exact)
2	1.253731	.3001409 7.516644	1.970588 (exact)
3	1.424779	.6240707 3.494844	5.539216 (exact)
Crude	1.263848	.6387976 2.631877	(exact)
M-H combined	1.305299	.6787116 2.510352	

Test of homogeneity (M-H)      chi2(2) =      0.56    Pr>chi2 = 0.7559

Test that combined OR = 1:

Mantel-Haenszel chi2(1) =      0.64

Pr>chi2 =      0.4241

This can be done also in one summary approach (but no crude ratios)

```
. tabodds Case_Control Age_of_mother [fweight = frequency], adjust(Age_of_Child) base(3)
```

Mantel-Haenszel odds ratios adjusted for Age\_of\_Child

Age_of_mother	Odds Ratio	chi2	P>chi2	[95% Conf. Interval]	
1	2.398149	2.87	0.0900	0.844416	6.810762
2	1.305299	0.64	0.4241	0.677862	2.513498
3	1.000000	.	.	.	.

Score test for trend of odds: chi2(1) = 2.61  
Pr>chi2 = 0.1059