

§ 流行病學原理：資料分析 bias (systemic error)

一. Crude analysis :

(一) Creating data

1. Data to create

	Case	Control
OC user	12	53
Non-user	30	347
Total	42	400

2. Code :

```
/* creating data */
data crude;
    input oc$ group$ count;
    datalines;
        user case 12
        user control 53
        non_user case 30
        non_user control 347
    ;
run;

proc print data = crude;
run;
```

3. Data created :

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of oc by group			
	oc	group		
		case	control	Total
	user	12	53	65
		2.71	11.99	14.71
		18.46	81.54	
		28.57	13.25	
	non_user	30	347	377
		6.79	78.51	85.29
		7.96	92.04	
		71.43	86.75	
	Total	42	400	442
		9.50	90.50	100.00

流行病學資料分析 –
SAS
Bias (Systemic error)
學號：b07401048
系級：醫學五

(二)Crude odd ratio

1. Result :

(1) Odds ratio 為 2.6189

(2) 95% CI : 1.2631-5.4300 , 達統計上顯著。

2. Figure :

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Odds Ratio	2.6189	1.2631	5.4300
Relative Risk (Column 1)	2.3200	1.2536	4.2935
Relative Risk (Column 2)	0.8859	0.7862	0.9982

3. Description :

(1) 分為 case 組（發生 venous thrombosis）與 control 組（沒有發生 venous thrombosis）。比較兩組曾服用 Oral-contraceptive（exposure）的比例。

(2) 將 case 組與 control 組服用 OC 的比例相除為 odd ratio。Odd ratio 為 2.6189，95% CI：1.2631-5.4300，達統計上顯著。

(3) 也就是說，發生 venous thrombosis 的病人暴露於 OC 的比例顯著高於沒有發生 venous thrombosis 的人，且為 2.6189 倍。

(4) 在 sampling rate independent of exposure 的前提下，exposure odd ratio 等於 risk ratio。也就是說，在抽樣與暴露相互獨立的前提下，服用 oral-contraceptive 的人與沒有服用 oral-contraceptive 的人發生 venous thrombosis 的 risk ratio 是 2.6189。

(5) 此 odd ratio 沒有經過分層處理，故稱為 crude analysis。

4. Code :

```
/* q1 : crude analysis */
title "q1 crude analysis";

proc freq data = crude order = data;
    tables oc * group / chisq or;
    weight count;
run;
```

二. Stratified analysis :

(一)Creating data

1. Data to create

	Age 20-29			Age 30-39	
	Case	Control		Case	Control
OC user	10	39		2	14
Non-user	18	158		12	189
Total	28	197		14	203

2. Code :

```
/* q2 : stratified analysis */
title "q2 stratified analysis";
/* creating data */
data stratified;
  input age$ oc$ group$ count;
  datalines;
    20_29 user case 10
    20_29 user control 39
    20_29 non_user case 18
    20_29 non_user control 158
    30_39 user case 2
    30_39 user control 14
    30_39 non_user case 12
    30_39 non_user control 189
  ;
run;
proc print data = stratified;
run;
```

3. Data created :

The FREQ Procedure

age=20_29

Frequency Percent Row Pct Col Pct	Table of oc by group			
	oc	group		
		case	control	Total
user		10	39	49
		4.44	17.33	21.78
		20.41	79.59	
		35.71	19.80	
non_user		18	158	176
		8.00	70.22	78.22
		10.23	89.77	
		64.29	80.20	
Total		28	197	225
		12.44	87.56	100.00

The FREQ Procedure

age=30_39

Frequency Percent Row Pct Col Pct	Table of oc by group			
	oc	group		
		case	control	Total
user		2	14	16
		0.92	6.45	7.37
		12.50	87.50	
		14.29	6.90	
non_user		12	189	201
		5.53	87.10	92.63
		5.97	94.03	
		85.71	93.10	
Total		14	203	217
		6.45	93.55	100.00

(二) Stratum-specific-ORs

1. Result :

- (1) 20-29 歲組的 odd ratio 為 2.2507，95% CI：0.9632 – 5.2592，未達統計上顯著。
- (2) 30-39 歲組的 odd ratio 為 2.2500，95% CI：0.4578 – 11.0584，未達統計上顯著。
- (3) 以 Breslow-Day test 檢定兩 stratum-specific odd ratio 是否相等。檢定結果 p-value 為 0.9997，兩 stratum-specific odd ratio 無顯著差異。
- (4) 因此以 Mantel-Haenszel method 計算 pooled odd ratio。Common odd ratio 為 2.2506，95% CI：1.0630 – 4.7649，達統計上顯著。

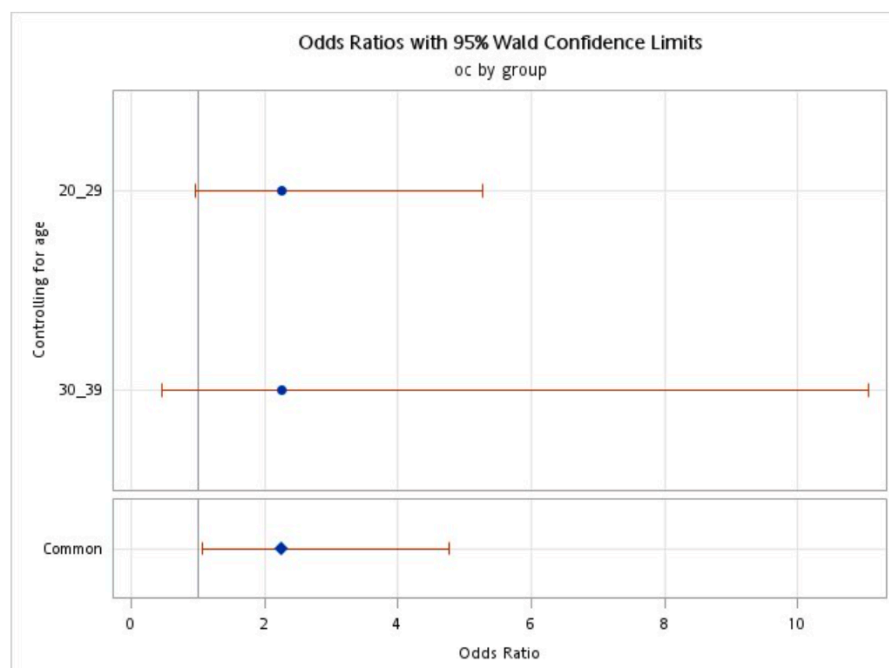
2. Figures :

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Odds Ratio	2.2507	0.9632	5.2593
Relative Risk (Column 1)	1.9955	0.9858	4.0394
Relative Risk (Column 2)	0.8866	0.7629	1.0304

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Odds Ratio	2.2500	0.4578	11.0594
Relative Risk (Column 1)	2.0938	0.5123	8.5563
Relative Risk (Column 2)	0.9306	0.7707	1.1235

Breslow-Day Test for Homogeneity of Odds Ratios	
Chi-Square	0.0000
DF	1
Pr > ChiSq	0.9997

Common Odds Ratio and Relative Risks				
Statistic	Method	Value	95% Confidence Limits	
Odds Ratio	Mantel-Haenszel	2.2506	1.0630	4.7649
	Logit	2.2506	1.0642	4.7596
Relative Risk (Column 1)	Mantel-Haenszel	2.0136	1.0715	3.7838
	Logit	2.0148	1.0725	3.7850
Relative Risk (Column 2)	Mantel-Haenszel	0.8993	0.7978	1.0136
	Logit	0.9034	0.8033	1.0161



3. Code :

```
/* stratum-specific odd ratio */  
title "q2-a";  
  
proc freq data = stratified order = data;  
  tables oc * group / chisq or cmh;  
  weight count;  
  by age;  
run;  
  
proc freq data = stratified order = data;  
  tables age * oc * group / chisq or cmh ;  
  weight count;  
run;
```

(三) Comparing stratum-specific ORs and crude ORs

1. 如第一題之結論，crude odd ratio 為 2.6189
2. 如上題之結論，adjusted odd ratio 為 2.2506。
3. 方向：away from NULL，導致 overestimate the effect of exposure。



(四) Whether age is a confounder for the association between OC use and venous thrombosis ?

1. Confounding effect :

$$\frac{(cOR - aOR)}{cOR} = \frac{2.6189 - 2.2506}{2.6189} = 14\%$$

2. 以 10% rule 判斷，則 age 符合 confounding 的定義。

三. OC use in source population among controls :

(一)Creating data

1. Data to create

	OC use	
	+	-
Age 20-29	39	158
Age 30-39	14	189

2. Code :

```
/* creating data */
data cf_e;
  input age$ oc$ count;
  datalines;
    20_29 user 39
    20_29 non_user 158
    30_39 user 14
    30_39 non_user 189
  ;
run;

proc print data = cf_e;
run;
```

3. Data created :

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of age by oc			
	age	oc		
		user	non_user	Total
20_29		39	158	197
		9.75	39.50	49.25
		19.80	80.20	
		73.58	45.53	
30_39		14	189	203
		3.50	47.25	50.75
		6.90	93.10	
		26.42	54.47	
Total		53	347	400
		13.25	86.75	100.00

(二) Odd ratio :

1. Result :

- (1) 以是否曾使用 OC 分組，分為 user 組與 non-user 組，比較兩組年齡於 20-29 歲及 30-39 歲的比例。
- (2) Odd ratio 為 3.3323，95% CI：1.7463 – 6.3587，達統計上顯著。
- (3) 也就是說，在抽樣與暴露無關的前提下，20-29 歲的人使用 OC 的 risk 是 30-39 歲的 3.3323 倍，且具顯著差異。

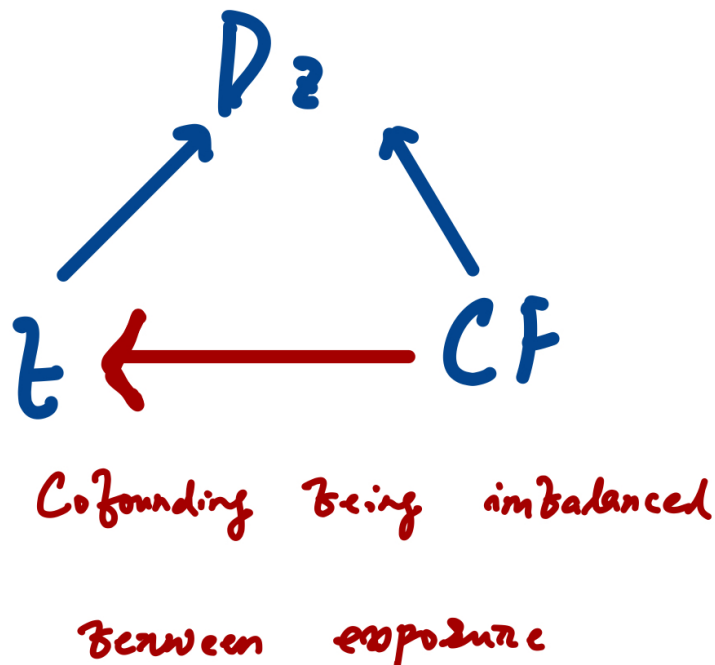
2. Figures :

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Odds Ratio	3.3323	1.7463	6.3587
Relative Risk (Column 1)	2.8706	1.6099	5.1184
Relative Risk (Column 2)	0.8614	0.7961	0.9321

3. Code :

```
/*q3 (a) odd ratio */  
proc freq data = cf_e order = data;  
  tables age * oc / chisq or;  
  weight count;  
run;
```

(三) Draw the triangle to show what does this OR indicate



(四) Which property of CF does this OR indicate

Ans : confounding must be imbalanced between exposure。

四. Age as an independent risk factor

(一) Creating data

1. Data to create

	Case	Control
Age 20-29	28	197
Age 30-39	14	203

2. Code :

```
/* q4 : cf_dz */
title "q4 cf_dz ";

/* creating data */
data cf_dz;
input age$ group$ count;
datalines;
    20_29 case 28
    20_29 control 197
    30_39 case 14
    30_39 control 203
;
run;

proc print data = cf_dz;
run;
```

3. Data created :

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of age by group			
	age	group		
		case	control	Total
	20_29	28	197	225
		6.33	44.57	50.90
		12.44	87.56	
		66.67	49.25	
	30_39	14	203	217
		3.17	45.93	49.10
		6.45	93.55	
		33.33	50.75	
	Total	42	400	442
		9.50	90.50	100.00

(二)Odd ratio :

1. Result :

- (1) 兩組 odd ratio 為 2.0609，95% CI：1.0537 – 4.0308，達統計上顯著。
- (2) 若符合抽樣與年齡相互獨立的前提，則 20-29 歲發生 venous thrombosis 的 rate 是 30-39 歲發生 venous thrombosis 的 rate 的 2.0609 倍，且達統計上顯著。
- (3) 代表 age 與 venous thrombosis 有 association，可能是 cause 或是 a proxy for a cause。

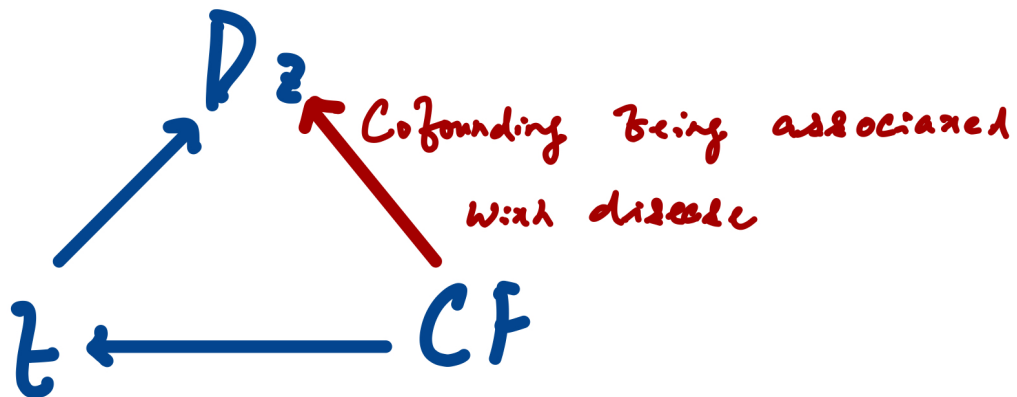
2. Figures :

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Odds Ratio	2.0609	1.0537	4.0308
Relative Risk (Column 1)	1.9289	1.0440	3.5637
Relative Risk (Column 2)	0.9359	0.8811	0.9942

3. Code :

```
/* odd ratio */  
proc freq data = cf_dz order = data;  
  tables age * group / chisq or;  
  weight count;  
run;
```

(三)What does this OR indicate in the triangle ?



(四)Which property of CF does this OR indicate ?

Ans : confounding must be associated with disease。