

世代研究的描述性統計 (1)

醫學研究部 醫學研究部 生統小組

徐倩儀

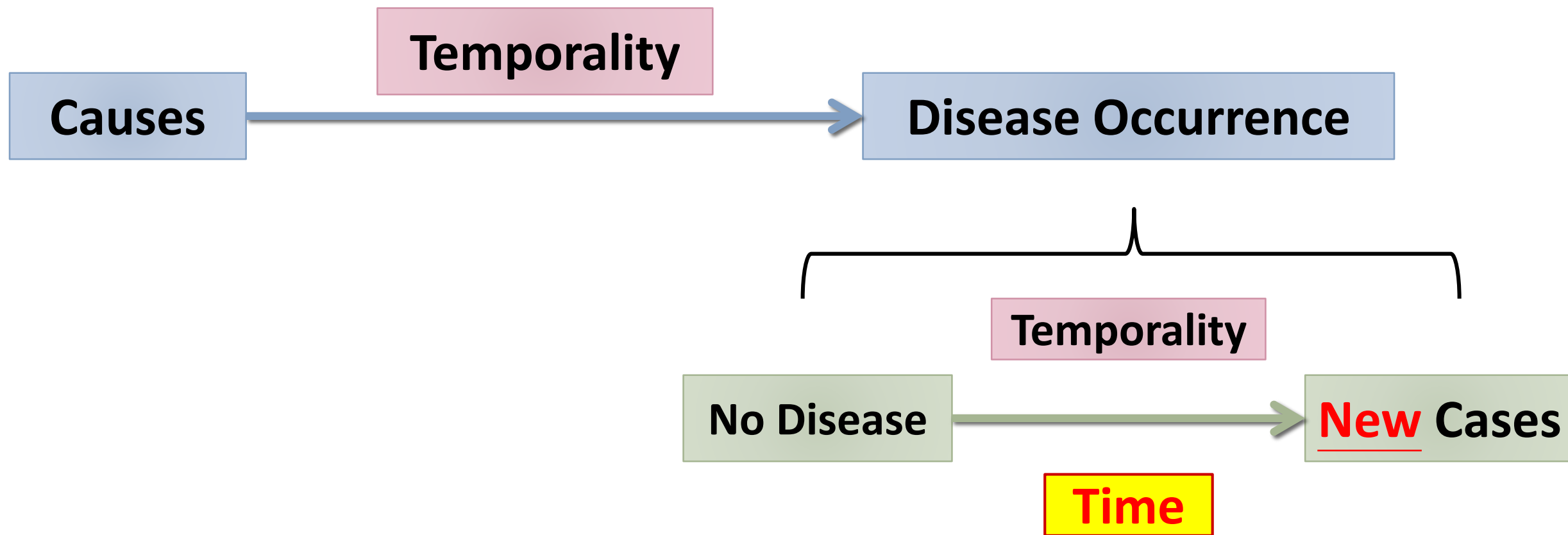
2023/11/1

內容大綱

- 如何製作「表1」
- 列聯表分析(卡方檢定、危險對比值odds ratio)
- 計算疾病的頻率(發生率)

世代研究：探討疾病的發生率及風險

判定因果關係的必要條件之一



定性資料

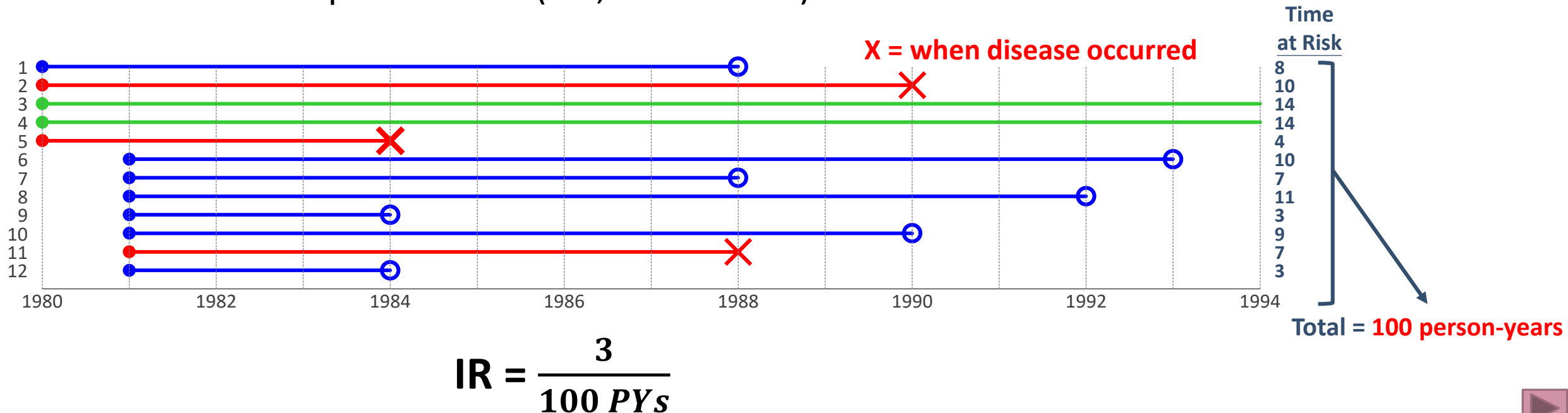
- 研究主題(流病)
 - 疾病、傷害、殘障、死亡
 - 兩分變項 (Binary variables)
- 頻率量數 (frequency measure)
 - 率、比率 (Rate)
 - 分率 (Proportion)
 - 比值 (Ratio)

Rates

- Time into the denominator
 - speed (kilometers per hour)
 - 120 kilometers in 2 hours
 - average speed : 60 km/hr.
- Health outcomes
 - "rates" are actually proportions
 - mortality rate / attack rate / case-fatality rate

Rates

- Like a true rate
 - incidence rates or incidence density
 - the number of health outcomes
 - the total exposure time (i.e., time at risk)



Incidence rate 專業期刊會如何呈現?

TABLE 2 | Incidence rates and effect sizes of outcomes by valve replacement status in ESRD group.

Outcomes	Variables	Total numbers	Event (%) / per 1,000 person-years	Models	Hazard ratios (95% CI)	P value
Total mortality	Patients with mechanical valve	456	412 (90.4%) / 457.4	0	1 (reference)	NA
				1		
				2		
				3		
	Patients with bioprosthetic valve	456	406 (89.0%) / 426.8	0	1.00 (0.95–1.06)	0.88
				1	0.98 (0.93–1.05)	0.55
				2	0.99 (0.93–1.05)	0.64
				3	0.88 (0.82–0.93)	<0.001
CV death	Patients with mechanical valve	456	236 (51.8%) / 262.0	0	1 (reference)	NA
				1		
				2		
				3		
	Patients with bioprosthetic valve	456	208 (45.6%) / 218.7	0	0.98 (0.90–1.06)	0.58
				1	0.96 (0.89–1.04)	0.29
				2	0.97 (0.89–1.05)	0.38
				3	0.83 (0.76–0.90)	<0.001

CI, confidence interval; CV, cardiovascular; ESRD, end-stage renal disease; NA, not available.

Model 0: crude effect size by the two groups.

Model 1: adjusted effect by age, sex.

Model 2: adjusted effect by age, sex, total number of valves replaced, hypertension, diabetes mellitus, congestive heart failure, coronary artery diseases, and chronic obstructive pulmonary disease.

Model 3: adjusted effect by age, sex, total number of valves replaced, hypertension, diabetes mellitus, congestive heart failure, coronary artery diseases, chronic obstructive pulmonary disease, and medications (antiarrhythmic agents of Ia Ib, Ic, III, calcium channel blockers, angiotensin receptor blockers, statins, insulin, oral hypoglycemic agents).

Incidence rate

如何計算?

$$\text{失智症的發生率} = \frac{\text{失智症的發生數}}{\text{總追蹤人年}} * 1000$$

(per 1000 person-years)

$$\text{Dementia (rate)} = \frac{\text{Dementia}}{\text{F/U dementia year}} * 1000$$

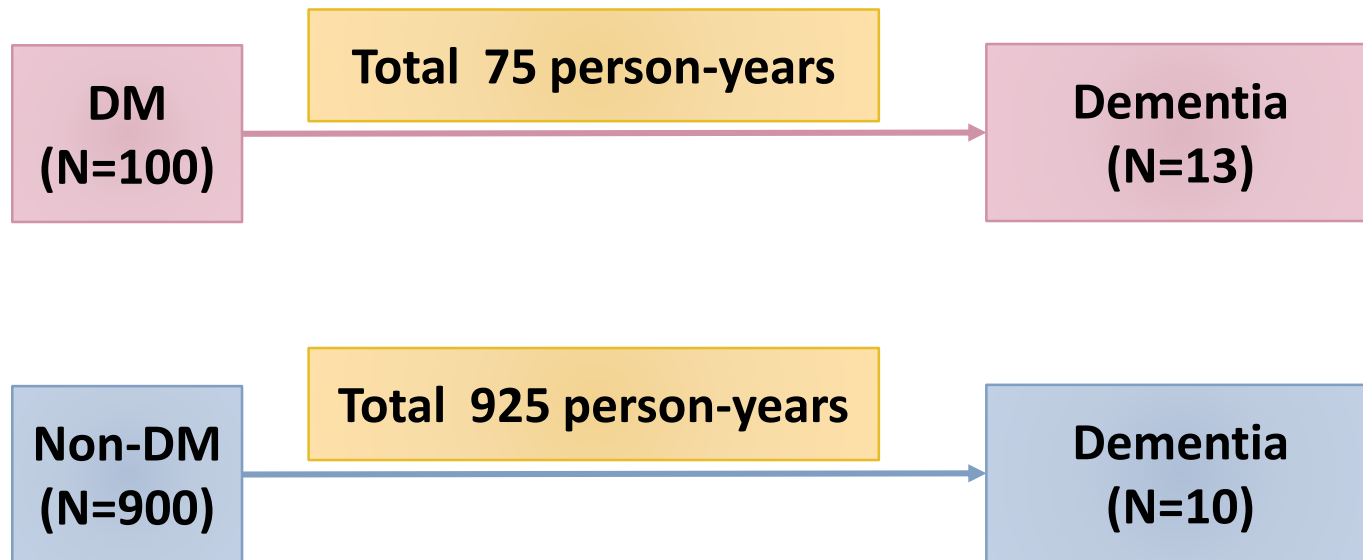
$$\text{發生率的95\% CI (Lower)} = A \times \text{Lower} \left[\text{dementia} - (\sqrt{\text{dementia}} * 1.96) \right]$$

$$\text{發生率的95\% CI (Upper)} = A \times \text{Upper} \left[\text{dementia} + (\sqrt{\text{dementia}} * 1.96) \right]$$

$$\frac{1000}{\text{F/U dementia year (總追蹤人年)}}$$

動手算算看 (Incidence rate)

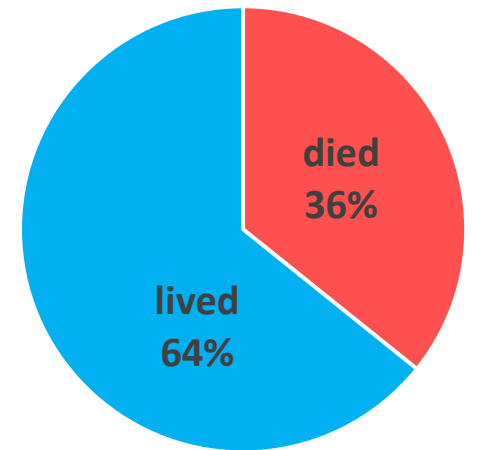
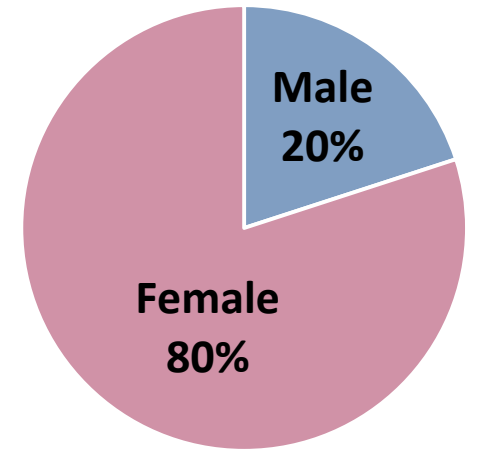
- Excel計算



課程範例I-2.xlsx
【發生率計算】

Proportions

- Relates a part to a whole
- Total class size is 100 (20 male / 80 female)
 - The proportions of
 - Male is $20/100$ or 20%
 - Female is $80/100$ or 80%
- 44 died / 79 lived
 - Proportion (died) : $44/(44+79) = 0.36$ or 36%
 - “case-fatality” rate



Prevalence (proportions)

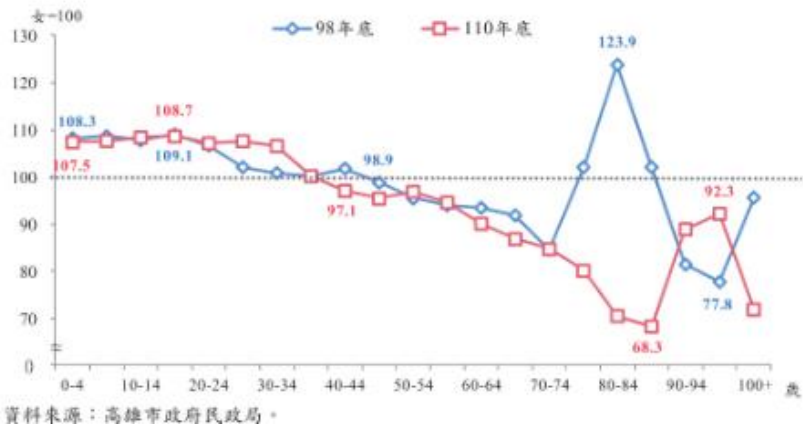
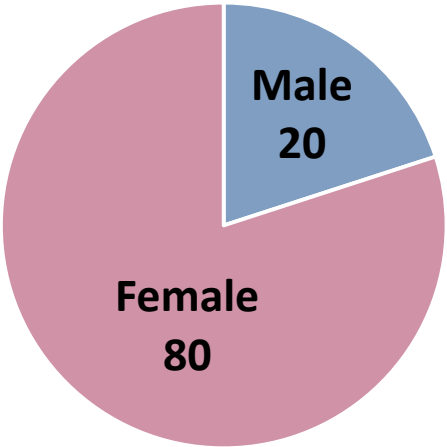
$$\text{Period prevalence} = \frac{\text{Number of cases that existed in a given period}}{\text{Number of people in the population during this period}}$$

(proportion)

- 點盛行率(期盛行率)
 - 某個時間點(或期間)，患某病的所有病例數佔全人口數的比例
 - 呈現方式
 - 百分比(%)或分數
 - 每一萬人或每十萬人患病的人數
- Prevalence is low (<10%)
 - Prevalence = incidence × duration

Simple Ratios

- A class
 - 20 male students and 80 female students
- The ratio of men to women
 - 20:80 (20/80)
 - 1:4 ratio (1/4 ratio)
 - 0.25



性比例=(男性人口數÷女性人口數)×100
意思：每一百個女性人口，有多少個男性人口？

年	男嬰	女嬰	性比例
一〇七年	93357	87299	106.9
一〇八年	90812	84262	107.8
一〇九年	83748	77540	108.0
一一〇年	81220	75799	107.2
一一一年	71208	66205	107.6
資料來源：內政部戶政司。			
更新日期：2023/6/10			

頻率量數 (frequency measure)

	率、比率 (Rate)	分率 (Proportion)	比值 (Ratio)
	兩個數量 <u>相除</u> 所得的值		
分母	可以包含 <u>時間</u>	可以包含 <u>時間</u>	<u>不</u> 包含時間
單位	<u>有單位</u> (人年數, 年 ⁻¹)	不一定有單位	無單位
定義	分子 < 分母	分子 < 分母	分子 ≠ 分母 (互斥)
範圍	零和無限大 ($0 \leq \text{Rate} \leq \infty$)	$0 \leq \text{Proportion} \leq 1$	無大小的限制
例子	發生率 Incidence rate 100 $\frac{\text{cases}}{\text{person-year}}$ 10,000 $\frac{\text{cases}}{\text{person-century}}$	盛行率、死亡率、致死率 Incidence proportion, Prevalence rate (proportion) Probability that disease develops within a specified time interval Incidence proportion = $\frac{\sum_{\text{persons}} \text{individual proportions}}{\text{initial size of the population}} = A/N$	比值、風險比 20:80 或 20/80 或 0.25



Risk calculations in a cohort study



		Disease develops	Disease dose not develop	Totals	Incidence Rates of Disease
First select	Exposed	<i>a</i>	<i>b</i>	<i>a</i> + <i>b</i>	$\frac{a}{a + b}$
	Not exposed	<i>c</i>	<i>d</i>	<i>c</i> + <i>d</i>	$\frac{c}{c + d}$
		$\frac{a}{a+b}$ = Incidence in exposed		$\frac{c}{c+d}$ = Incidence in non-exposed	

$$\text{Relative risk (RR)} = \frac{\text{Incidence in exposed}}{\text{Incidence in non-exposed}} = \frac{\left(\frac{a}{a+b}\right)}{\left(\frac{c}{c+d}\right)}$$

- RR=1, No association
- RR>1, Positive association (possibly causal)
- RR<1, Negative association (possibly protective)

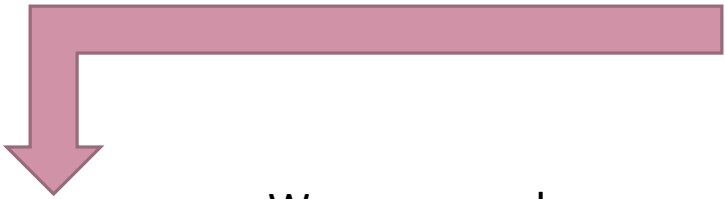
Risk calculations in a cohort study

Hypothetical cohort study, the relation of smoking to the development of coronary heart disease over a **1-year period**.

	CHD develops	CHD Dose not develop	Totals	Incidence per 1,000 per Year
Smoke	84	2,916	3,000	$\frac{84}{3000} * 1000 = 28.0$ per 1,000
Do not smoke	87	4,913	5,000	$\frac{87}{5000} * 1000 = 17.4$ per 1,000

$$\text{Relative risk (RR)} = \frac{\text{Incidence in exposed}}{\text{Incidence in non-exposed}} = \frac{28.0}{17.4} = 1.61$$

Case-Control study



		First Select		Odds
		Cases (With Disease)	Controls (Without Disease)	
Then Measure Past Exposure	Were exposed	a	b	$\frac{a}{b}$ (exposed)
	Were not exposed	c	d	$\frac{c}{d}$ (non-exposed)
Odds		$\frac{a}{c}$ (case, disease)	$\frac{b}{d}$ (control, without disease)	

$$\text{Odds ratio (OR)} = \frac{\text{Odds (exposed)}}{\text{Odds (non-exposed)}} = \frac{a/b}{c/d} = \frac{ad}{bc}$$

$$\text{Odds ratio (OR)} = \frac{\text{Odds (case,disease)}}{\text{Odds (control,without disease)}} = \frac{a/c}{b/d} = \frac{ad}{bc}$$

- OR=1, No association
- OR>1, Positive association (possibly causal)
- OR<1, Negative association (possibly protective)

Relative risk & Odds Ratio

	Develop Disease	Do Not Develop Disease	
Exposed	200	9800	10,000
Not Exposed	100	9900	10,000

罕見疾病

Relative risk (RR) $\approx$ Odds ratio (OR)

$$= \frac{200/10000}{100/10000}$$

$$= 2$$

$$= \frac{200 \cdot 9900}{100 \cdot 9800}$$

$$= 2.02$$

	Develop Disease	Do Not Develop Disease	
Exposed	50	50	100
Not Exposed	25	75	100

Relative risk (RR) $\neq$ Odds ratio (OR)

$$= \frac{50/100}{25/100}$$

$$= 2$$

$$= \frac{50 \cdot 75}{25 \cdot 50}$$

$$= 3$$

a	b
c	d

動手算算看 (Odds Ratio)

Case-Control Study

		First Select	
		CHD Cases	Controls
Past Exposure	Smokers	112	176
	Non smokers	88	224

Odds Ratio (OR) = ?

課程範例I-2.xlsx
【OR計算】

Odds Ratio

Unmatched study (10 cases and 10 controls)

	Cases	Control
Exposed	6	3
Non Exposed	4	7

$$\text{Odds ratio} = \frac{ad}{bc} = \frac{6*7}{3*4} = 3.5$$

Matched study (10 cases and 10 matched controls)

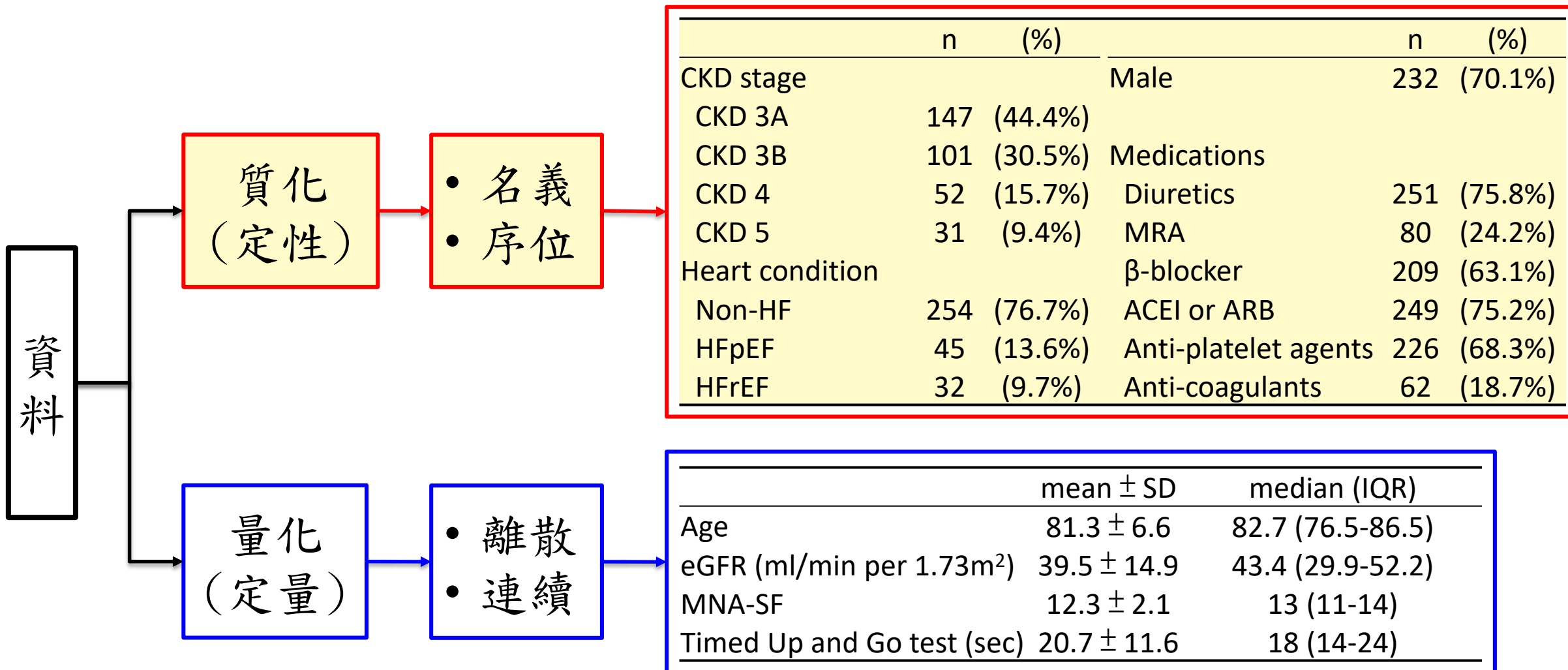
		Control	
		Exposed	Non Exposed
Case	Exposed	2	4
	Non Exposed	1	3

$$\text{Odds ratio (matched pairs)} = \frac{b}{c} = \frac{4}{1} = 4$$

a	b
c	d

資料分類

資料呈現



專業期刊中 Table 1如何呈現?

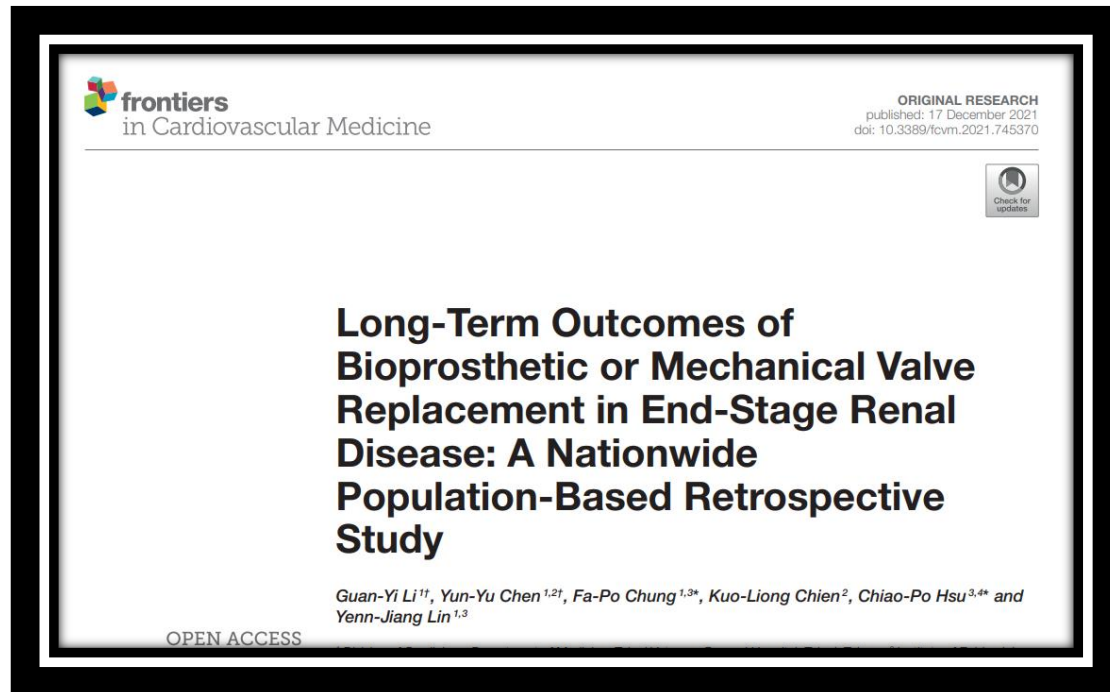


TABLE 1 | Baseline characteristics of ESRD cohorts after propensity-score matching.

Variables	ESRD group (N = 912)		P value
	Mechanical valve (N = 456)	Bioprosthetic valve (N = 456)	
Age	67.4 ± 11.8	66.8 ± 11.9	0.51
Male gender	240 (52.6%)	254 (55.7%)	0.35
Valve location			
Aortic valve	242 (53.1%)	234 (51.3%)	0.60
Mitral valve	246 (53.9%)	254 (55.7%)	0.60
Tricuspid valve	13 (2.9%)	27 (5.9%)	0.02
Pulmonary valve	0 (0%)	3 (0.7%)	0.24
Total number of valves replaced	1.10 ± 0.30	1.14 ± 0.35	0.08
1	411 (90.1%)	395 (86.6%)	0.18
2	45 (9.9%)	59 (12.9%)	
3	0 (0%)	2 (0.44%)	
4	0 (0%)	0 (0%)	
Comorbidities			
ESRD (%)	456 (100%)	456 (100%)	>0.99
Diabetes mellitus (%)	16 (3.5%)	22 (4.8%)	0.32
Hypertension (%)	53 (11.6%)	62 (13.6%)	0.37
COPD (%)	4 (0.9%)	4 (0.9%)	>0.99
CHF (%)	118 (25.9%)	110 (24.1%)	0.54
Prior stroke (%)	16 (3.5%)	22 (4.8%)	0.32

資料整理

- 列聯表
 - 各分類次數

NO	CKD	HC	Death	sex	Med1	Med2	...
0001	3	1	0	1	0	0	
0005	3	1	0	0	0	0	
0013	2	3	0	1	1	1	
0018	2	1	0	1	0	0	
0021	2	1	0	1	0	0	
0023	4	1	1	1	1	1	
0024	2	1	0	1	0	0	
0027	4	1	0	0	1	0	
0029	4	1	1	0	1	0	
0052	2	1	0	0	1	0	

CKD

CKD 3A CKD 3B CKD 4 CKD 5

Sex Male
Female

101 71 39 21
46 30 13 10

Chi-square (χ^2) test

- 獨立性檢定
 - 兩組類別變數是否獨立(是否有關聯)
- 適合度檢定
 - 分佈比例是否一致

Chi-square (χ^2) test

- 獨立性檢定
 - 兩組類別變數是否有關聯

觀察值(O)			
	有病	沒病	
抽菸	A	B	A+B
不菸	C	D	C+D
	A+C	B+D	N

期望值(E)		
	有病	沒病
抽菸	$\frac{(A+B)(A+C)}{N}$	$\frac{(A+B)(B+D)}{N}$
不菸	$\frac{(A+C)(C+D)}{N}$	$\frac{(B+D)(C+D)}{N}$

- 80%以上細格期望值>5
 - 小樣本或未達到
 - Fisher's Exact test

- 每一細格觀察值(O)偏離期望值(E)的程度 $= (O-E)/E$
- 避免負號(取平方)
- $\chi^2 = \sum (O-E)^2 / E$

練習範例

frontiers
in Cardiovascular Medicine

ORIGINAL RESEARCH
published: 04 June 2021
doi: 10.3389/fcvm.2021.680098

Check for updates

Impacts of Heart Failure and Physical Performance on Long-Term Mortality in Old Patients With Chronic Kidney Disease

Shuo-Chun Weng^{1,2}, Yu-Chi Chen³, Chiann-Yi Hsu⁴, Chu-Sheng Lin⁵, Der-Cherng Tarn^{1,6,7,8} and Shih-Yi Lin^{1,9*}

TABLE 1 | Baseline characteristics of older patients with different staging of CKD.

Characteristics	CKD stage 3A (n = 147)	CKD stage 3B (n = 101)	CKD stage 4 (n = 52)	CKD stage 5 (n = 31)	p-value
Age, years	83.1 (77.1–86.3)	81.8 (76.0–86.7)	83.2 (79.0–87.1)	78.8 (73.3–83.6)	0.043
Male	101 (68.7)	71 (70.3)	39 (75.0)	21 (67.7)	0.846
Heart condition					0.001
Non-heart failure	119 (81.0)	84 (83.2)	36 (69.2)	15 (48.4)	
HFpEF	15 (10.2)	12 (11.9)	10 (19.2)	8 (25.8)	
HFrEF	13 (8.8)	5 (4.9)	6 (11.5)	8 (25.8)	
Medications					
Diuretics	108 (73.5)	72 (71.3)	47 (90.4)	24 (77.4)	0.054
MRA	31 (21.1)	16 (15.8)	18 (34.6)	15 (48.4)	0.001
β-blocker	88 (59.9)	60 (59.4)	41 (78.9)	20 (64.5)	0.078
ACEI or ARB	110 (74.8)	77 (76.2)	43 (82.7)	19 (61.3)	0.183
Anti-platelet agents	95 (64.6)	66 (65.4)	44 (84.6)	21 (67.7)	0.052
Anti-coagulants	24 (16.3)	13 (12.9)	19 (36.5)	6 (19.4)	0.003

練習範例

- Chi-square (χ^2) test

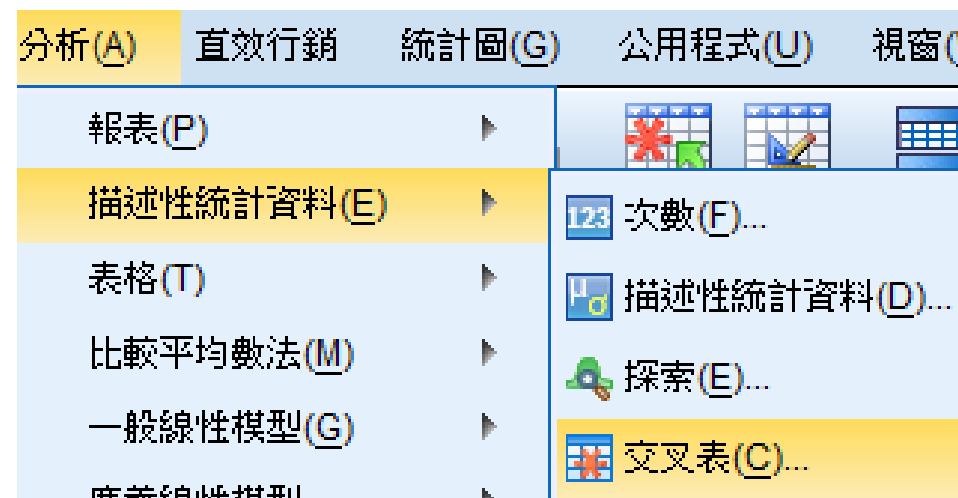
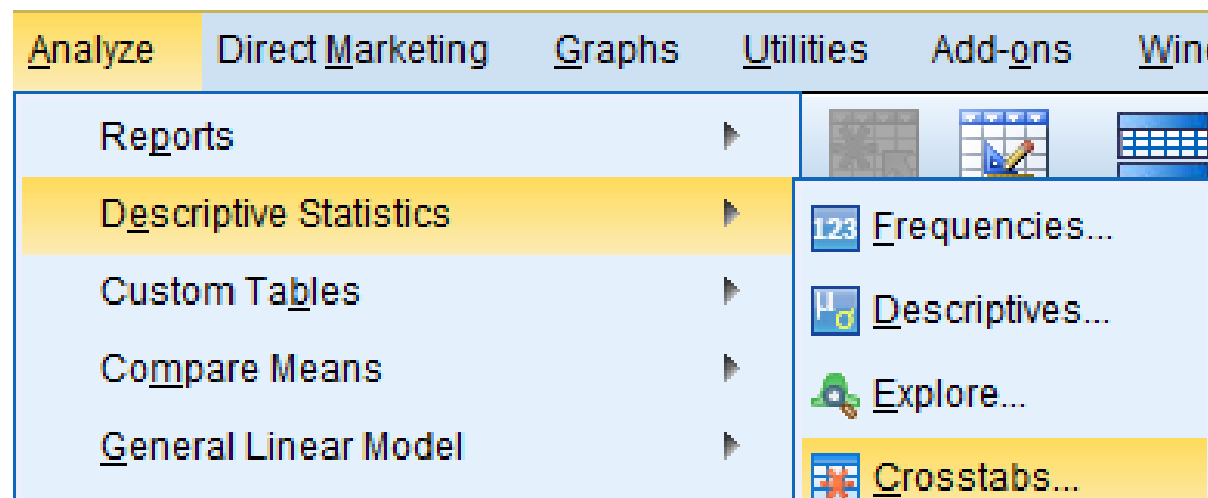
- Excel

- CHISQ.TEST()

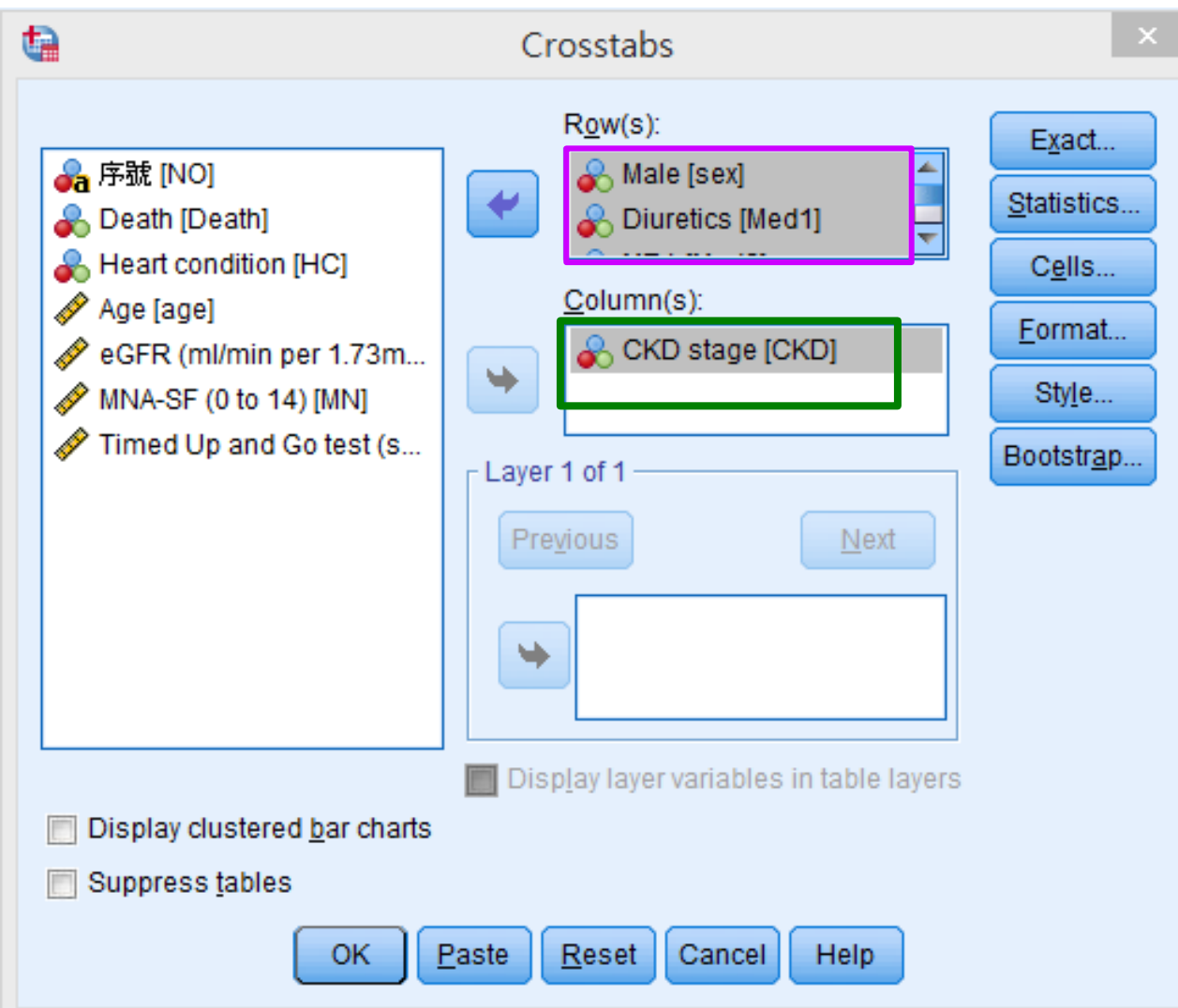
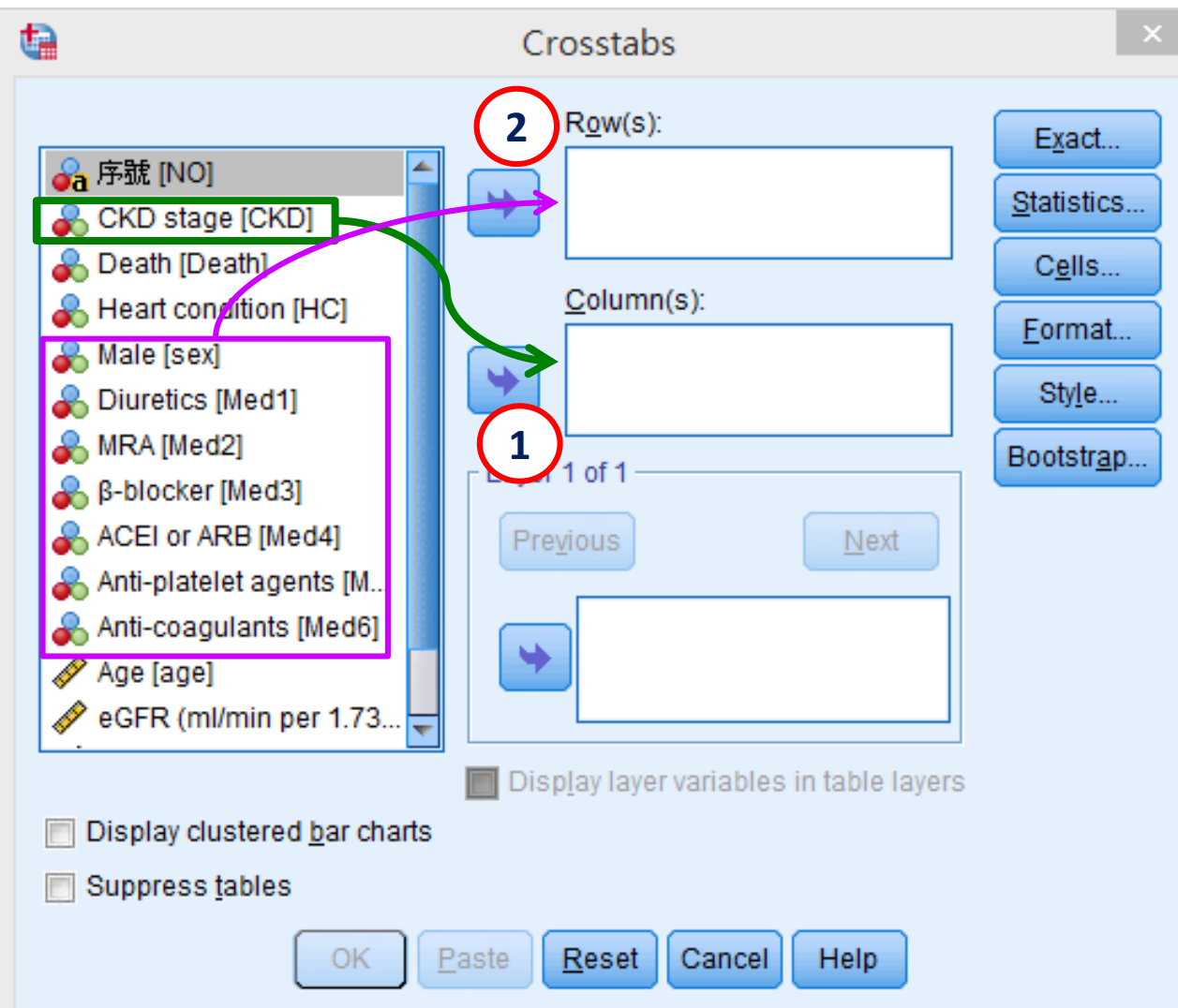
課程範例I-2.xlsx
【OR計算】

- SPSS

- Analyze → Descriptive Statistics → Crosstabs



練習範例-SPSS (Crosstabs)



練習範例-SPSS (Crosstabs)

The image displays two SPSS dialog boxes for a Crosstabs analysis, with annotations indicating specific settings.

Crosstabs: Statistics (Left Dialog, Purple Outline):

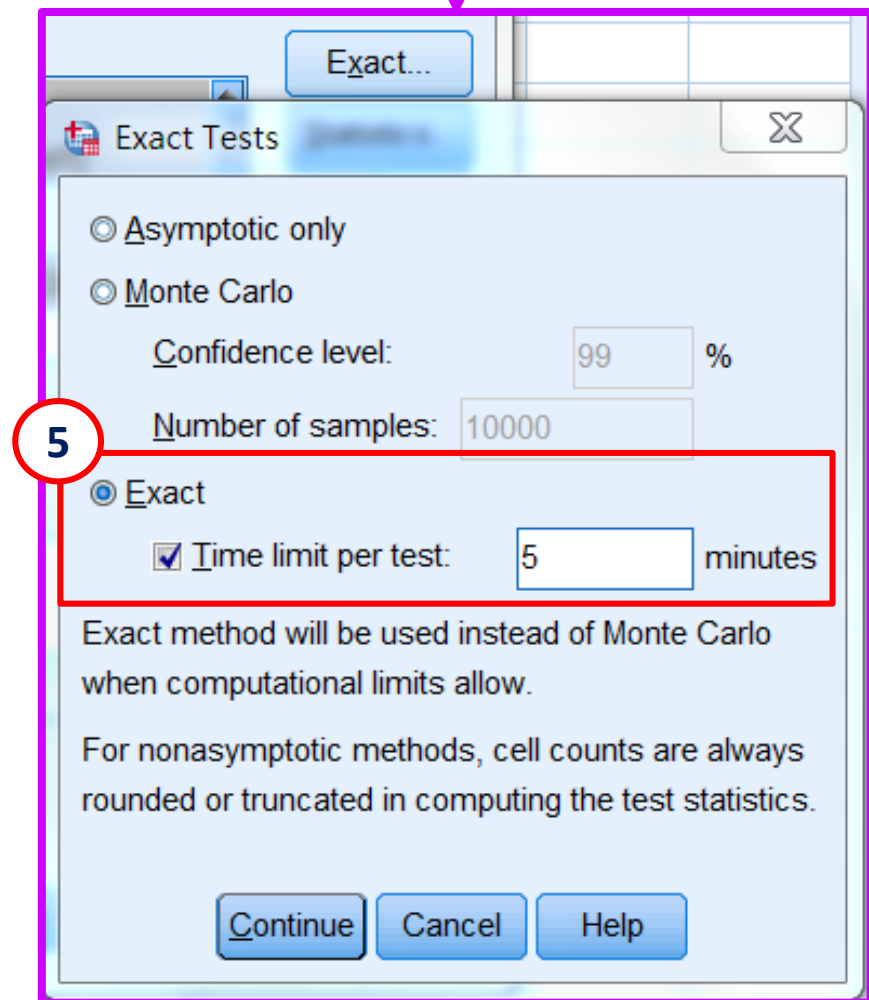
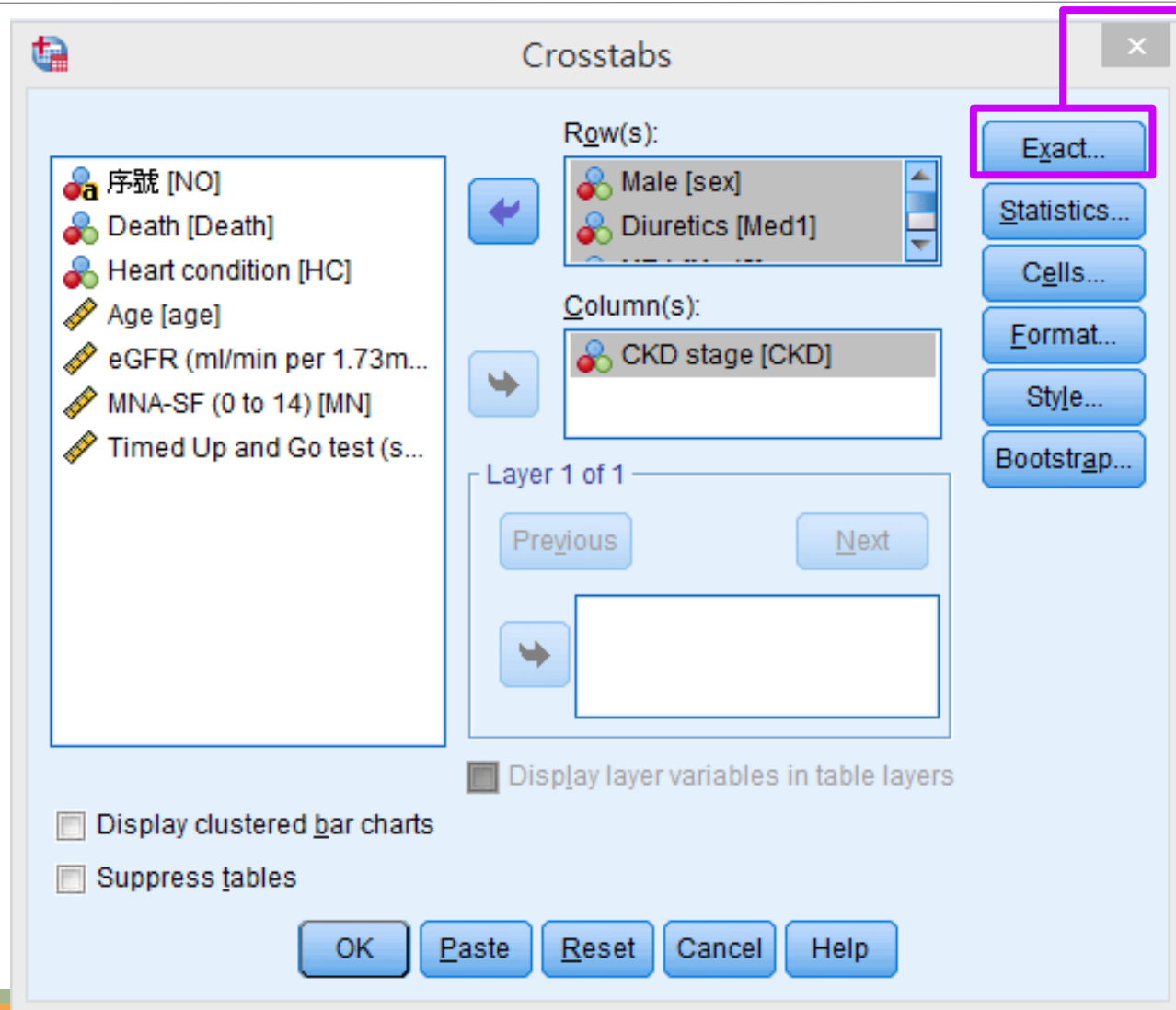
- 3** (Red circle) highlights the **Chi-square** checkbox, which is checked.
- The **Statistics...** button is highlighted with a green box.
- The **Cells...** button is highlighted with a green box.
- Other visible options include: Correlations, Nominal (Contingency coefficient, Phi and Cramer's V, Lambda, Uncertainty coefficient), Ordinal (Gamma, Somers' d, Kendall's tau-b, Kendall's tau-c), Nominal by Interval (Eta), Kappa, Risk, McNemar, and Cochran's and Mantel-Haenszel statistics.
- The test common odds ratio equals:
- Buttons: Continue, Cancel, Help.

Crosstabs: Cell Display (Right Dialog, Green Outline):

- 4** (Red circle) highlights the **Column** checkbox under the **Percentages** section, which is checked.
- Other visible options include: Counts (Observed, Expected, Hide small counts, Less than 5), z-test (Compare column proportions, Adjust p-values (Bonferroni method)), Residuals (Unstandardized, Standardized, Adjusted standardized), and Noninteger Weights (Round cell counts, Round case weights, Truncate cell counts, Truncate case weights, No adjustments).
- Buttons: Continue, Cancel, Help.

A green arrow points from the **Cells...** button in the Statistics dialog to the **Crosstabs: Cell Display** dialog.

練習範例-SPSS (Crosstabs)



練習範例-SPSS (Crosstabs)

Crosstab

			CKD stage				Total
			CKD 3A	CKD 3B	CKD 4	CKD 5	
Heart condition	Non-HF	Count	119	84	36	15	254
		% within CKD stage	81.0%	83.2%	69.2%	48.4%	76.7%
	HFpEF	Count	15	12	10	8	45
		% within CKD stage	10.2%	11.9%	19.2%	25.8%	13.6%
	HFrEF	Count	13	5	6	8	32
		% within CKD stage	8.8%	5.0%	11.5%	25.8%	9.7%
	Total	Count	147	101	52	31	331
		% within CKD stage	100.0%	100.0%	100.0%	100.0%	100.0%

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	21.558 ^a	6	.001
Likelihood Ratio	19.173	6	.004
Linear-by-Linear Association	11.965	1	.001
N of Valid Cases	331		

a. 2 cells (16.7%) have expected count less than 5. The minimum expected count is 3.00.

test * Death Crosstabulation

			Death		Total
			Alive	Death	
test	No	Count	30	4	34
		% within Death	10.5%	9.1%	10.3%
	Yes	Count	257	40	297
		% within Death	89.5%	90.9%	89.7%
Total		Count	287	44	331
		% within Death	100.0%	100.0%	100.0%

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	.077 ^a	1	.782	1.000	.517
Continuity Correction ^b	.000	1	.992		
Likelihood Ratio	.079	1	.778		
Fisher's Exact Test					
Linear-by-Linear Association	.077	1	.782		
N of Valid Cases	331				

a. 1 cells (25.0%) have expected count less than 5. The minimum expected count is 4.52.

練習範例-SPSS (Crosstabs-OR)

Row(s):
Diuretics [Med1]

Column(s):
Death [Death]

1 of 1

Previous Next

Display layer variables in table layer

Reset Cancel Help

Exact... Statistics...

Crosstabs: Statistics

☐ Chi-square ☐ Correlations

Nominal

☐ Contingency coefficient
☐ Phi and Cramer's V
☐ Lambda
☐ Uncertainty coefficient

Ordinal

☐ Gamma
☐ Somers' d
☐ Kendall's tau-b
☐ Kendall's tau-c

Nominal by Interval

☐ Eta ☐ Kappa ☒ Risk ☐ McNemar

☐ Cochran's and Mantel-Haenszel statistics

Test common odds ratio equals: 1

Continue Cancel Help

	age	eGFR	MN	TU
	84	41	14	15
	72	53	12	13

		Death		Total
		Death	Alive	
Diuretics	Yes	43	208	251
	No	1	79	80
Total		44	287	331

a	b
c	d

Risk Estimate

	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for Diuretics (Yes / No)	16.332	2.211	120.612
For cohort Death = Death	13.705	1.918	97.941
For cohort Death = Alive	.839	.789	.892
N of Valid Cases	331		

問卷調查

Thank you



For your attention!!