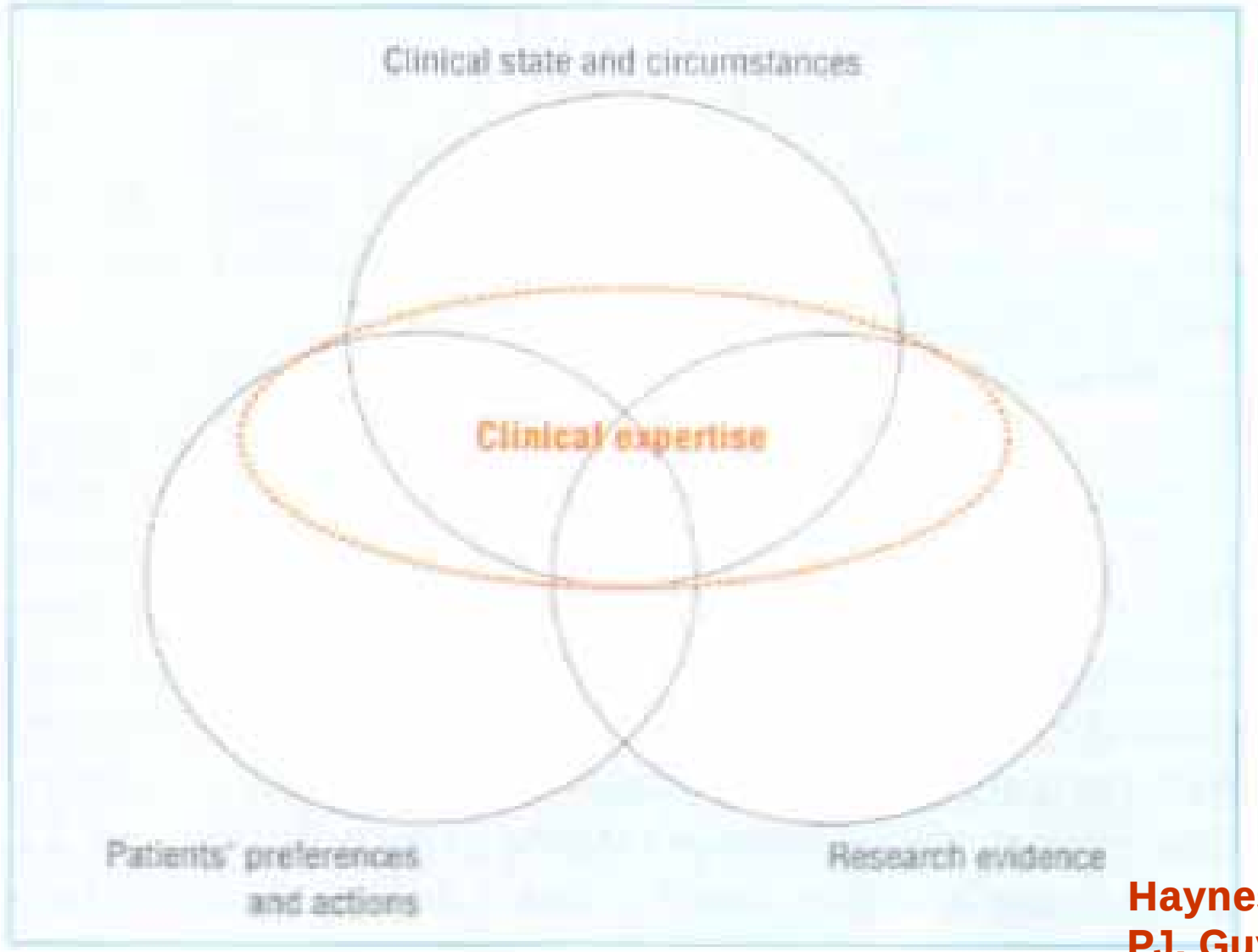
如何評閱醫學文獻
(How to Critically Appraising Evidence)

(EBM基礎課程for PGY1)

光田醫院大甲分院
家庭醫學科
賴文恩　醫師

實証醫學(Evidence Based Medicine)



An updated model for evidence based clinical decisions[1]

Haynes RB, Deveaux PJ, Guyatt GH. BMJ 2002; 324, 7350

Five Steps to Practice EBM

- Step 1 asking a question
- Step 2 searching for the best evidence
- Step 3 critically appraising
- Step 4 applying
- Step 5 evaluating

Five Types of Question

- Diagnosing and screening
- Therapy
- Harm/etiology
- Prognosis
- Guidelines

Diagnosis and Screening

Clinical Scenario

- 張先生，48歲男性，由於
  被告知 C 型肝炎帶原 [anti-HCV(+)]。
  醫師建議以後，
  應每半年追蹤 肝臟超音波/胎兒蛋白 檢查

- 結果，第一次追蹤就發現，雖然
  肝臟超音波正常，但AFP高達45.6ng/mL

- 他再來門診時，醫師說...

The Effectiveness of Serum α-Fetoprotein Level in Anti-HCV Positive Patients for Screening Hepatocellular Carcinoma

Yen-Chun Peng
Chi-Sen Chan
Gran-Hum Chen

Division of Gastroenterology, Department of Internal Medicine, Taichung Veterans General Hospital, Taichung, Taiwan

Corresponding Author:
Chi-Sen Chang, MD
Division of Gastroenterology
Department of Internal Medicine
Taichung Veterans General Hospital
160, Sec 3, Chung-Kang Rd
Taichung 407
Taiwan
Tel: +886-4-3741331
Fax: +886-4-3741318

KEY WORDS: α-fetoprotein; HCV; Hepatocellular carcinoma
ABBREVIATIONS: Hepatitis B Virus (HBV); Hepatitis C Virus (HCV); Hepatocellular Carcinoma (HCC)

ABSTRACT

BACKGROUND/AIMS: In Taiwan, most cases of hepatocellular carcinoma (HCC) are hepatitis B virus (HBV) or hepatitis C virus (HCV) related. The serum α-fetoprotein (AFP) level is an important factor in the diagnosis of HCC. There have been many studies discussing the role of AFP in diagnosing HBV-related HCC, but only few concerning HCV-related HCC. In this study, we aimed at analyzing the distribution of AFP levels in anti-HCV positive patients with and without HCC and evaluating the effectiveness of serum AFP levels in screening HCV-related HCC.

METHODOLOGY: From 1993-1996, we collected the AFP data of 205 HCC patients retrospectively, who were anti-HCV positive. For comparison, 131 randomized anti-HCV positive patients without evidence of HCC served as the control group. We analyzed the AFP distribution in both groups over the following ranges: ≤5ng/ml, >5-20ng/ml, >20-50ng/ml, >50-100ng/ml, >100-200ng/ml and >200-400ng/ml, and >400ng/ml.

RESULTS: The distributions of AFP levels in anti-HCV positive patients with HCC were 13.2%, 21.5%, 11.2%, 4.9%, 4.4%, 7.3%, and 37.6%. The distributions in anti-HCV positive patients without evidence of HCC were 34.3%, 55.0%, 8.4%, 1.5%, 0.8%, 0%, 0%.

CONCLUSIONS: We found the differences in AFP to be statistically significant between anti-HCV positive patients with and without HCC. A serum AFP level of more than 200ng/ml highly indicates HCC. However, there is a large overlap between these 2 groups. Thus, in anti-HCV positive patients, AFP level is not a good single reference for diagnosis of HCC. Anti-HCV positive patients should be routinely screened for HCC by image studies along with serum AFP level.

METHODOLOGY

Between January 1993 and December 1996, 205 HCC patients (167 males, 38 females) with anti-HCV positive and negative for hepatitis B surface antigen (HBsAg) in our hospital were enrolled into this study. The age ranged from 35-85 years (mean: 65.0±8.9 years). HCC was diagnosed by ultrasonography or CT scan findings. Definite diagnosis was made by liver biopsy or a specific vascular lesion by highly selective celiac angiography. The AFP levels of these patients at the time of definite diagnosis of HCC were recorded. For comparison, 131 patients with anti-HCV positive and HBsAg negative, without HCC, were randomly selected. During the period of follow-up, AFP and abdominal ultrasonography were performed, every 3-6 months, to screen for HCC. Liver CT scan or hepatic angiography was ordered for patients with suspicious liver lesions on ultrasonography. If HCC was proven, patients were transferred to the HCC group.

Critically Appraising Diagnostic test (VIP)

- (Valid)

  Evidence about a diagnostic test valid ?

- (Important)

  How important the evidence is?

  Accuracy of the test to distinguish p’ts with or without disorder

- (aPply)

  Can I apply this valid, accurate test to a specific patient?

Evidence about a Diagnostic Test **Valid**?

- An independent, blind comparison with golden standard of diagnosis?
  - Pt undergone both test in question & reference standard.
  - Results should be blinded to personnel of the other side
  - Avoid the conscious and unconscious bias (over-interpreted, or under–interpreted)
- Reference standard universally applied
  - when the reference standard is invasive or risky, sufficiently prolonged follow-up is OK
- Evaluated in an appropriate spectrum of patients *(like those we would use it in practice)*?
- Validated in a 2nd, independent groups of patients

How important the Evidence Is?

- Accuracy of Diagnostic test
  - Sensitivity / Specificity
  - Positive predictive value (PPV)
    / Negative predictive value (NPV)
  - Likelihood Ratio + / Likelihood Ratio -

例：某次乳癌社區篩檢 5000 位婦女，事後經 黃金標準 檢驗 發現，真正有病的 100 位，有 80 位檢測陽性，沒病的 4900 位中也有 200 位陽性：

	有病 Pr(D+)	無病 Pr(D-)	
Test (+)	a	b	a + b
Test (−)	c	d	c + d
	a + c	b + d	a+b+c+d

	D(+)	D(-)	
T(+)	**80**	**200**	**280**
->T(-)	**20**	**4700**	**4720**
	100	**4900**	**5000**

- Sensitivity = a/(a+c) = 80/100 = 0.8
- Specificity = d/(b+d) = 4700/4900 = 0.96
- Positive predictive value
  = a/(a+b) = 80/280 = 0.286
- Negative predictive value
  = d/(c+d) = 4700/4720 = 0.996
- Prevalence
  = (a+c)/(a+b+c+d) = 100/5000 = 0.02

	有病 Pr(D+)	無病 Pr(D-)	
Test (+)	a	b	a + b
Test (—)	c	d	c + d
	a + c	b + d	a+b+c+d

	D(+)	D(-)	
T(+)	**80**	**200**	**280**
->T(-)	**20**	**4700**	**4720**
	100	**4900**	**5000**

	Pr(D+)	Pr(D-)
T+	a/(a+b+c+d)	b/(a+b+c+d)
T -	c/(a+b+c+d)	d/(a+b+c+d)
	(a+c)/(a+b+c+d)	(b+d)/(a+b+c+d)

	D(+)	D(-)	
T(+)	**0.016**	**0.04**	**0.056**
—>T(-)	**0.004**	**0.94**	**0.944**
	0.02	**0.98**	**1**

	Pr(D+)	Pr(D-)
T+	a/(a+c) x (a+c)/(a+b+c+d)	b/(b+d) x (b+d)/(a+b+c+d)
T -	c/(a+c) x (a+c)/ (a+b+c+d)	d/(b+d) x (b+d)/ (a+b+c+d)
	(a+c)/(a+b+c+d)	(b+d)/(a+b+c+d)

	Pr(D+)	Pr(D-)
T+	Sen x Prev	(1-Spe) x (1-Prev)
T -	(1-Sen) x Prev	Spe x (1-Prev)
	Prev	(1-Prev)

SpPin and SnNout

- SpPin
  - Extremely high (**Sp**)ecificity,
    a (**P**)ositive result tends to Rule (**in**) the diagnosis.

	$Pr(D^+)$	$Pr(D^-)$
T+	Sen x Prev	~0
T -	(1-Sen) x Prev	1-Prev
	Prev	(1-Prev)

- SnNout
  - Extremely high (**Sen**)sitivity,
    a (**N**)egative result tends to rule (**out**) the diagnosis.

	$Pr(D^+)$	$Pr(D^-)$
T+	Prev	(1-Spe) x (1-Prev)
T -	~0	Spe x (1-Prev)
	Prev	(1-Prev)

	有病 Pr(D+)	無病 Pr(D-)	
Test (+)	a	b	a + b
Test (—)	c	d	c + d
	a + c	b + d	a+b+c+d

	D(+)	D(-)	
T(+)	**80**	**200**	**280**
->T(-)	**20**	**4700**	**4720**
	100	**4900**	**5000**

	$Pr(D^+)$	$Pr(D^-)$
T+	a/(a+b+c+d)	b/(a+b+c+d)
T -	c/(a+b+c+d)	d/(a+b+c+d)
	(a+c)/(a+b+c+d)	(b+d)/(a+b+c+d)

	D(+)	D(-)	
T(+)	**0.016**	**0.04**	**0.056**
->T(-)	**0.004**	**0.94**	**0.944**
	0.02	**0.98**	**1**

	$Pr(D^+)$	$Pr(D^-)$
T+	a/(a+c) x (a+c)/(a+b+c+d)	b/(b+d) x (b+d)/(a+b+c+d)
T -	c/(a+c) x (a+c)/ (a+b+c+d)	d/(b+d) x (b+d)/ (a+b+c+d)
	(a+c)/(a+b+c+d)	(b+d)/(a+b+c+d)

	$Pr(D^+)$	$Pr(D^-)$
T+	prev x sen	(1-prev)x(1-spe)
T -	prev x(1-sen)	(1-prev)x spe
	Pr(D)=prev	1-Pr(D)=1-prev

x(Sen) *x(1-Spe)*

假設對某一種檢查而言，Sen, Spe 爲固定，

其 PPV & NPV 將隨 Pr(D)而有很大的不同。

例如：某一種檢查 sensitivity = 0.85, specificity = 0.9

	$Pr(D^+)$	$Pr(D^-)$
T+	0.001 x 0.85	0.999 x 0.1
T -	0.001 x 0.15	0.999 x 0.9
	0.001	0.999

In population A with **[Pr(D) = 0.1%]**：

$$PPV = \frac{0.001 \times 0.85}{0.001 \times 0.85 + 0.999 \times 0.1} = 0.84\%$$

$$NPV = \frac{0.999 \times 0.9}{0.001 \times 0.15 + 0.999 \times 0.9} = 99.98\%$$

	$Pr(D^+)$	$Pr(D^-)$
T+	0.05 x 0.85	0.95 x 0.1
T -	0.05 x 0.15	0.95 x 0.9
	0.05	0.95

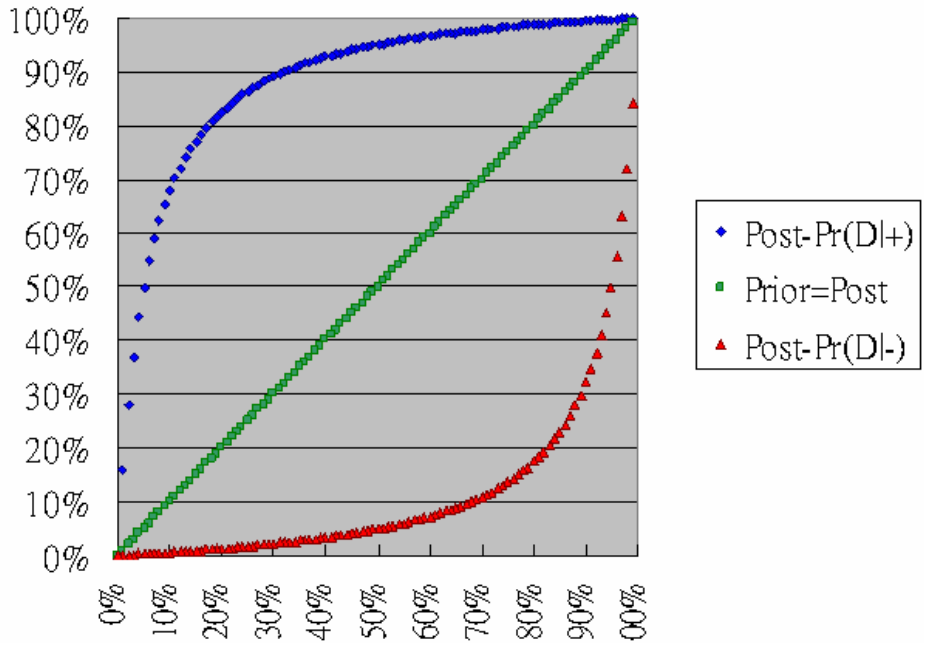
In population B with **[Pr(D) = 5%]**：

$$PPV = \frac{0.05 \times 0.85}{0.05 \times 0.85 + 0.95 \times 0.1} = 30.91\%$$

$$NPV = \frac{0.95 \times 0.9}{0.05 \times 0.15 + 0.95 \times 0.9} = 99.13\%$$

如果有一個檢查的 Sen = 0.95, Spe = 0.95，
那麼盛行率對陽性預測值與陰性預測值的影響如下：

Prevalence Prior Pr(D)	99%	95%	90%	80%	70%	60%	50%	40%	30%	20%	10%	5%	1%	0.5%	0.1%
PPV= Post-Pr(D|+)	99.9%	99.7%	99.4%	98.7%	97.8%	96.6%	95.0%	92.7%	89.1%	82.6%	67.9%	50.0%	16.1%	8.7%	1.9%
NPV	16.1%	50.0%	67.9%	82.6%	89.1%	92.7%	95.0%	96.6%	97.8%	98.7%	99.4%	99.7%	99.9%	99.97%	99.99%
1-NPV= Post-Pr(D|-)	83.9%	50.0%	32.1%	17.4%	10.9%	7.3%	5.0%	3.4%	2.2%	1.3%	0.6%	0.3%	0.1%	0.03%	0.01%



判讀檢查結果，應考慮「盛行率 (事前機率)」

如果判讀檢查結果時，沒有考慮 Pr(D)=Prev.，可能導致誤差。

例一：

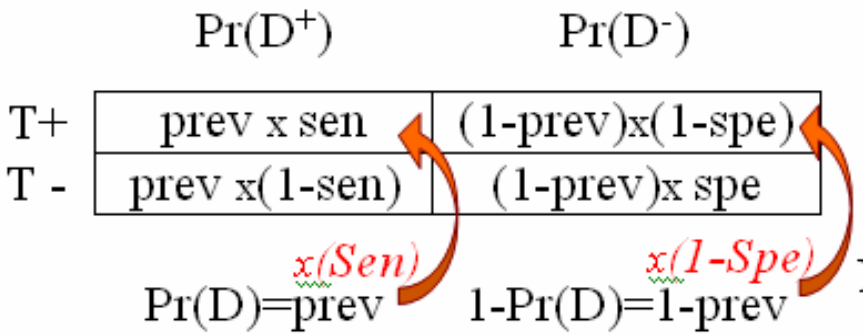
1) 在山地鄉，Chronic cough patient with CXR(+)，T.B.可能性較大；
2) 在都市，Chronic cough patient with CXR(+)， T.B.可能性較小；
3) 在美國，Chronic cough patient with CXR(+)， T.B.可能性更小。

例二：

如果最近報載，東海校園發現 Dengue Fever，那麼，兩位症狀完全相同，抽血檢驗也都+的病人，醫師向病人解釋患病的可能性時，仍然受到與東海地緣關係的影響。

例三：

Lower back pain 症狀完全一模一樣的 patient, 出現在家醫科門診與免疫風濕科門診的疾病可能性 Pr(D) 不同，Test 的判讀也不同。



$$PPV=a/(a+b)=\frac{\text{prev x sen}}{\text{prev x sen}+(1-\text{prev})(1-\text{spe})}$$

$$NPV=d/(c+d)=\frac{(1-\text{prev})\text{ x spe}}{\text{prevx}(1-\text{sen})+(1-\text{prev})\text{xspe}}$$

因此，Pr(D)也可叫**事前機率**；而 Pr(D|+)或 Pr(D|-)叫做**事後機率**。

意即：檢查後，依檢查結果，將得病的**機率**，**Revise** 成爲事後機率。

$$PPV=a/(a+b)=\frac{\text{Prev x sen}}{\text{Prevxsen}+(1-\text{Prev})(1-\text{spe})}=\frac{\mathbf{Pr(D)}\text{ x sen}}{\mathbf{Pr(D)}\text{ x sen}+(1-\mathbf{Pr(D)})(1-\text{spe})}$$

$$NPV=d/(c+d)=\frac{(1-\text{Prev})\text{ x spe}}{\text{Prevx}(1-\text{sen})+(1-\text{Prev})\text{xspe}}=\frac{(1-\mathbf{Pr(D)})\text{ x spe}}{\mathbf{Pr(D)}\text{x}(1-\text{sen})+(1-\mathbf{Pr(D)})\text{xspe}}$$

以勝算(odds)爲基礎的計算

然而，前式太複雜，不易理解應用。我們改用 Odds 的概念來取代 Pr(D)，會得到一個比較簡潔的公式

$$\text{Odds} = \frac{\text{有病的機率}}{\text{沒病的機率}} = \frac{\Pr(D)}{1-\Pr(D)}$$

	$\Pr(D^+)$	$\Pr(D^-)$
T+	prev x sen	(1-prev)x(1-spe)
T -	prev x (1-sen)	(1-prev) x spe

x(Sen) x(1-Spe)

$\Pr(D)$=prev 1-$\Pr(D)$=1-prev

$$\text{Pre-Odds} = \Pr(D^+)/(1-\Pr(D^-)) = \text{prev}/(1-\text{prev})$$

$$\text{Post-Odds}_{T+} = \frac{\text{prev} \times \text{sen}}{(1-\text{prev}) \times (1-\text{spe})} = \text{Pre-odds} \times LR_+$$

$$\text{Post-Odds}_{T-} = \frac{\text{prev} \times (1-\text{sen})}{(1-\text{prev}) \times \text{spe}} = \text{Pre-odds} \times LR_-$$

$$\text{Likelihood Ratio of "+" } (LR_+) = \frac{\text{Sensitivity}}{1-\text{Specificity}}$$

$$\text{Likelihood Ratio of "-" } (LR_-) = \frac{1-\text{Sensitivity}}{\text{Specificity}}$$

Posterior odds = LR_+ x Pre-odds if Test (+)

Posterior odds = LR_- x Pre-odds if Test(-)

Real time usage of Diagnostic test with EXCEL®

	D+	D-	Likelyhood Ratio (LR)
T+	SEN	1-SPE	= SEN / 1-SPE
T-	1-SEN	SPE	= 1-SEN / SPE
	1	1	

	D+	D-	Likelyhood Ratio (LR)
T+	0.700	0.180	= 0.7 / 0.18 = 3.89
T-	0.300	0.820	= 0.3 / 0.82 = 0.37
	1	1	

Post (T+) Odds = Pre Odds * LR(+)
Post (T-) Odds = Pre Odds * LR(-)

Odds = Pr(D) / Pr(D-)

= Pr(D) / (1-Pr(D))

Pr(D) = Odds/(1+Odds)

Pre-Test Odds * LR = Post-Test Odds

Pre-Test Pr(D)	Pre-Test Odds	LR(+)	Post (T+) Odds	Post (T+) Pr(D)
0.100	0.1/0.9=0.111	SEN/(1-SPE) = 3.89	0.111*3.89= 0.432	0.432/(0.432+1)=0.302

Pre-Test Pr(D)	Pre-Test Odds	LR(-)	Post (T-) Odds	Post (T-) Pr(D)
0.100	0.1/0.9=0.111	(1-SEN)/SPE = 0.37	0.111*0.37= 0.041	0.041/(0.041+1)=0.039

Multi-level Likelihood Ratio

Table 3.8 The usefulness of five levels of a diagnostic test result

Diagnostic test result	Serum ferritin (mmol/L)	Target disorder (Iron deficiency) present		Target disorder absent		Likelihood ratio	Diagnostic impact
		Number	%	Number	%		
Very positive	< 15	474	59 (474/809)	20	1.1 (20/1770)	52	Rule-in "SpPin"
Moderately positive	15–34	175	22 (175/809)	79	4.5 (79/1770)	4.8	Intermediate high
Neutral	35–64	82	10 (82/809)	171	10 (171/1770)	1	Indeterminate
Moderately negative	65–94	30	3.7 (30/809)	168	9.5 (168/1770)	0.39	Intermediate low
Extremely negative	≥ 95	48	5.9 (48/809)	1332	75 (1332/1770)	0.08	Rule-out "SnNout"
		809	100 (809/809)	1770	100 (1770/1770)		

Multi-level Likelihood Ratio

	D+	D-	Likelyhood Ratio (LR)
T+	SEN	1-SPE	= SEN / 1-SPE
T-	1-SEN	SPE	= 1-SEN / SPE
	1	1	

ferritin	Iron Deficiency(+)	Iron Deficiency(-)	Likelyhood Ratio (LR)
<15	58.6%	1.1%	= 0.586 / 0.011 = 51.85
15~34	21.6%	4.5%	= 0.216 / 0.045 = 4.85
35~64	10.1%	9.7%	= 0.101 / 0.097 = 1.05
65~94	3.7%	9.5%	= 0.037 / 0.095 = 0.39
>95	5.9%	75.3%	= 0.059 / 0.753 = 0.08
	1	1	

Post-Test Odds = Pre-Test Odds * LR

Odds = Pr(D) / Pr(D-)
= Pr(D) / (1-Pr(D))

Pr(D) = Odds/(1+Odds)

Pre-Test Odds * LR = Post-Test Odds

Pre-Test Pr(D)	Pre-Test Odds	LR	Post-Test Odds	Post-Test Pr(D)
0.100	0.1/0.9=0.111	4.85	0.111*4.85 = 0.538	0.538/(0.538+1)=0.35

1/10 -> 1/9 x 4.85 = 4.85/9 -> 4.85/13.85=0.35

(≒5/14 = 0.357)

Likelihood Ratio of common test or signs or symptoms

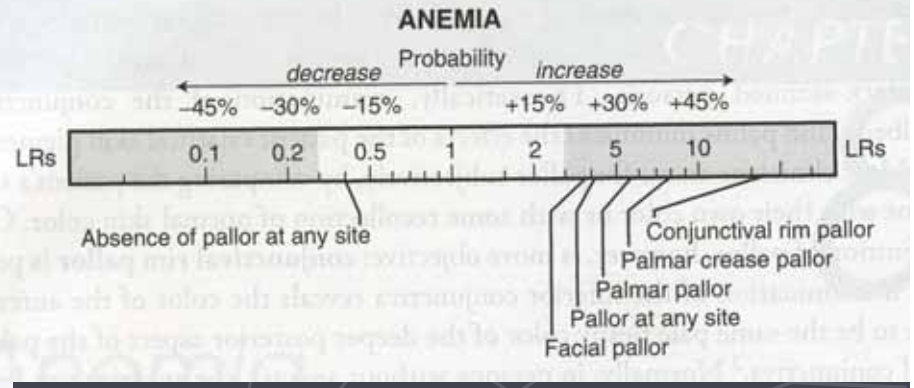
Box 8-1 Anemia*

Finding (Ref)†	Sensitivity (%)	Specificity (%)	Likelihood Ratio if Finding	
			Present	Absent
Pallor at any site[3–5]	38–77	66–92	**4.1**	0.4
Facial pallor[4]	46	88	**3.8**	0.6
Nail bed pallor[4,5]	59–60	66–93	NS	0.5
Palmar pallor[4,5]	58–64	74–96	**5.6**	0.4
Palmar crease pallor[4]	8	99	**7.9**	NS
Conjunctival pallor[4–7]	31–62	82–97	**4.7**	0.6
Conjunctival rim pallor[2]				
Pallor present	10	99	**16.7**	...
Pallor borderline	36	...	2.3	...
Pallor absent	53	16	0.6	...

NS, not significant; likelihood ratio (LR) if finding present = positive LR; LR if finding absent = negative LR.

**Diagnostic standard: For anemia, hematocrit <35%,[4] hemoglobin <11 g/dL,[2,5-7] or hemoglobin <11 g/dL in women and <13 g/dL in men.[3]*

†Definition of findings: For pallor at any site, examination of skin, nailbeds, and conjunctiva[3-5]; for facial pallor, the study excluded black patients; for palmar crease pallor, examination after gentle extension of the patient's fingers; for conjunctival rim pallor, see text.



Likelihood Ratio of common test or signs or symptoms

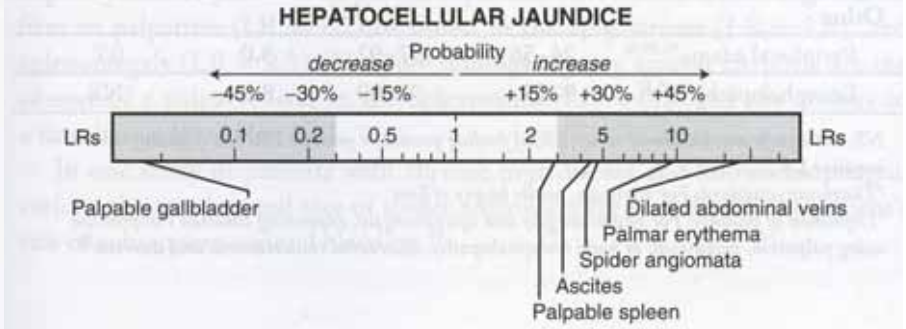
Box 6-1 Findings Predicting Hepatocellular Jaundice in Patients with Jaundice*

Finding (Ref)†	Sensitivity (%)	Specificity (%)	Likelihood Ratio if Finding	
			Present	Absent
General appearance				
Weight loss[31,33]	10–49	21–97	NS	NS
Skin				
Spider angiomata[31,33]	35–47	88–97	**4.7**	0.6
Palmar erythema[31]	49	95	**9.8**	0.5
Dilated abdominal veins[31]	42	98	**17.5**	0.6
Abdomen				
Ascites[31]	44	90	**4.4**	0.6
Palpable spleen[31,33]	29–47	83–90	2.9	0.7
Palpable gallbladder[31]	0†	69	**0.04**	1.4
Palpable liver[31,33]	71–83	15–17	NS	NS
Liver tenderness[31,33]	37–38	70–78	NS	NS

NS, not significant; likelihood ratio (LR) if finding present = positive LR; LR if finding absent = negative LR.

**Diagnostic standard: For* nonobstructive (vs. obstructive) jaundice, *needle biopsy of liver, surgical exploration, or autopsy.*

†None of the 41 patients with medical jaundice in this study had a palpable gallbladder; for calculation of the LRs, 0.5 was added to all cells of the 2 × 2 table.



E. RENAL COLIC

In one study of 1333 patients presenting with acute abdominal pain, two findings were accurate signs of ureterolithiasis (as diagnosed by imaging or follow-up): loin tenderness (sensitivity 15%, specificity 99%, positive LR = 27.7, negative LR = 0.9) and renal tenderness (sensitivity 86%, specificity 76%, positive LR = 3.6, negative LR = 0.2). As compelling as these findings are, they are less important than the finding of microscopic hematuria, which has a sensitivity of 75%, specificity of 99%, positive LR of 73.1, and negative LR of 0.3.[71]

Likelihood Ratio of common test or signs or symptoms

Box 48-1 Acute Abdominal Pain, Signs Detecting Peritonitis*

Finding (Ref)†	Sensitivity (%)	Specificity (%)	Likelihood Ratio if Finding Present	Likelihood Ratio if Finding Absent
Abdominal examination				
Guarding[2,26–33]	13–76	56–97	2.6	0.6
Rigidity[2,30–32,34]	6–40	86–100	**3.9**	NS
Rebound tenderness[2,26–40]	40–95	20–89	2.1	0.5
Percussion tenderness[33]	65	73	2.4	0.5
Abnormal bowel sounds[2,32]	25–61	44–95	NS	0.8
Rectal examination				
Rectal tenderness[2,29,30,32,33,35,36,41]	20–53	41–96	NS	NS
Other tests				
Positive abdominal wall tenderness test[16,42]	1–5	32–72	**0.1**	NS
Positive cough test[14,26,34,40]	73–84	44–79	1.8	0.4

NS, not significant; likelihood ratio (LR) if finding present = positive LR; LR if finding absent = negative LR.

**Diagnostic standard: For peritonitis, surgical exploration and follow-up of patients not operated on; causes of peritonitis included appendicitis (most common), cholecystitis, and perforated ulcer. One study also included patients with pancreatitis.[32]*

†Definition of findings: For abnormal bowel sounds, absent, diminished, or hyperactive; for abdominal wall tenderness test, see text; for positive cough test, the patient is asked to cough, and during the cough shows signs of pain or clearly reduces the intensity of the cough to avoid pain.[26]

Box 48-2 Acute Right Lower Quadrant Tenderness, Signs Detecting Appendicitis*

Finding (Ref)†	Sensitivity (%)	Specificity (%)	Likelihood Ratio if Finding Present	Likelihood Ratio if Finding Absent
Vital signs				
Fever[26,36,39,44]	47–81	40–70	1.5	0.6
Abdominal examination				
Severe right lower quadrant tenderness[26,27]	87–99	8–65	NS	**0.2**
McBurney's point tenderness[26,27,45]	50–94	75–86	**3.4**	0.4
Rovsing's sign[27,28,31,41]	22–68	58–96	2.5	0.7
Rectal examination				
Rectal tenderness[29,30,33,35,36,41]	38–53	41–62	NS	NS
Other signs				
Psoas sign[28,29,33]	13–42	79–97	2.0	NS
Obturator sign[29]	8	94	NS	NS

NS, not significant; likelihood ratio (LR) if finding present = positive LR; LR if finding absent = negative LR.

**Diagnostic standard: For appendicitis, surgical findings, histology, and follow-up of patients not operated on.*

†Definition of findings: For fever, temperature > 37.3° C[36,39,44] or not defined[26]; for positive cough test, see EBM Box 48-1.

重覆接受不同的檢查，即反覆 Revise 得病機率至 acceptable 的地步。

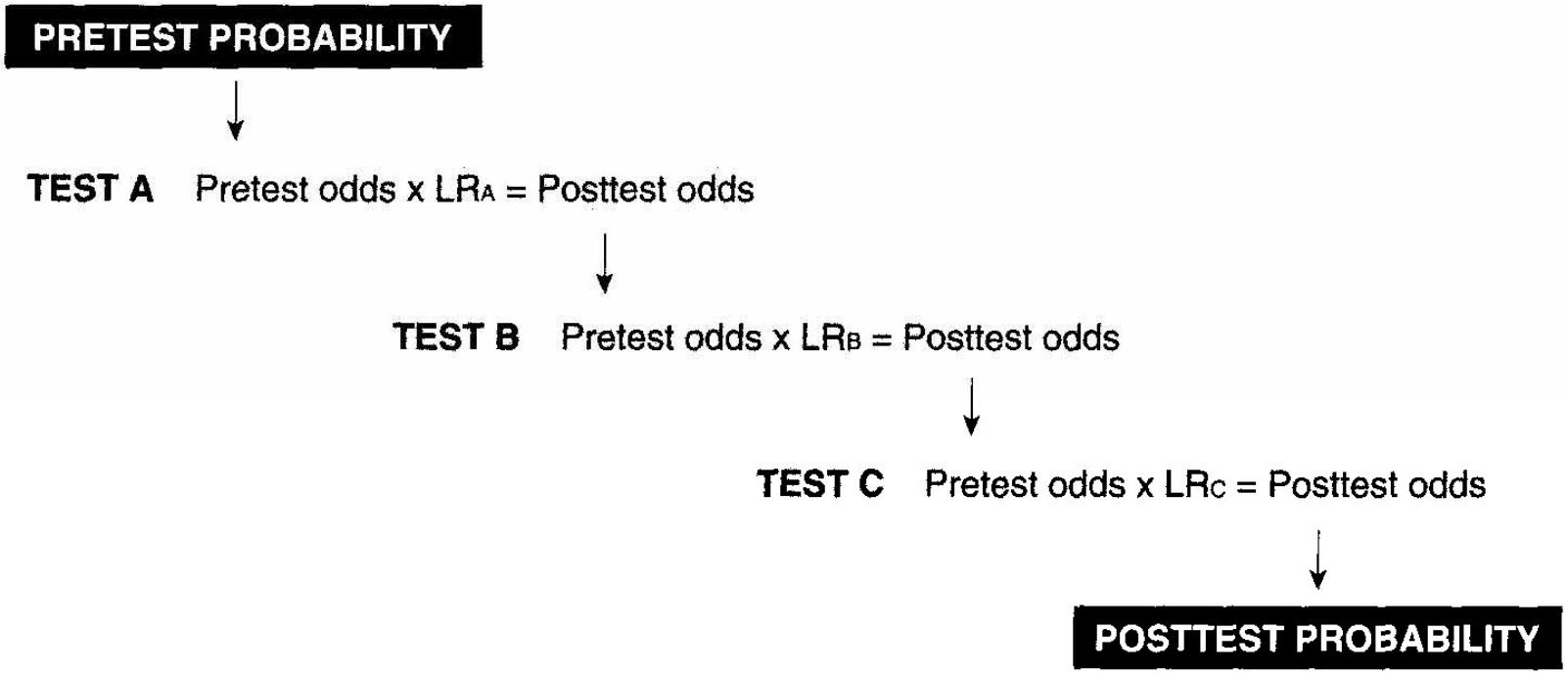


FIGURE 3.13 Use of likelihood ratios in serial testing. As each test is completed, its posttest odds become the pretest odds for the subsequent test.

TABLE 2 The Distribution of AFP in Both Groups

	HCC (+), n=205	HCC (-), n=131
AFP ≤5ng/ml	27 (13.2%)	45 (34.3%)
5<AFP≤20ng/ml	44 (21.5%)	45 (34.3%)
20<AFP≤50ng/ml	23 (11.2%)	11 (8.4%)
50<AFP≤100ng/ml	10 (4.9%)	2 (5.2%)
100<AFP≤200ng/ml	9 (4.4%)	1 (0.8%)
200<AFP≤400ng/ml	15 (7.3%)	0
AFP>400ng/ml	77 (37.6%)	0

HCC (+): anti-HCV positive patients with evidence of HCC;
HCC (-): anti-HCV positive patients without evidence of HCC;
AFP: α-fetoprotein; n: patient number

	HCC (+), n=205	HCC (-), n=131
AFP ≤5ng/ml	27 (13.2%)	45 (34.3%)
5<AFP≤20ng/ml	44 (21.5%)	45 (34.3%)
20<AFP≤50ng/ml	23 (11.2%)	11 (8.4%)
50<AFP≤100ng/ml	10 (4.9%)	2 (5.2%)
100<AFP≤200ng/ml	9 (4.4%)	1 (0.8%)
200<AFP≤400ng/ml	15 (7.3%)	0
AFP>400ng/ml	77 (37.6%)	0

anti HCV (+)	HCC(+)	HCC(-)	Likelihood Ratio (LR)	
AFP ≦ 5	0.132	0.343	0.132 / 0.343 =	0.385
5 < AFP≦ 20	0.215	0.55	0.215 / 0.55 =	0.391
20 < AFP≦ 50	0.112	0.084	0.112 / 0.084 =	1.333
50 < AFP≦ 100	0.049	0.015	0.049 / 0.015 =	3.267
100 < AFP≦ 200	0.044	0.008	0.044 / 0.008 =	5.500
200 < AFP≦ 400	0.073	0	0.073 / 0 =	∞
400 < AFP	0.376	0	0.376 / 0 =	∞

Odds = Pr(D) / Pr(D-)
= Pr(D) / (1-Pr(D)) → **Pr(D) = Odds/(1+Odds)**

Pre-Test Odds * LR = Post-Test Odds

Test1

Pre-Test Pr(D)	Pre-Test Odds	LR	Post-Test Odds	Post-Test Pr(D)
0.020	0.02/0.98=0.02	1.333	0.027	0.027/(0.027+1) = 0.026

Can I apply this test to a specific patient?

- Is the diagnostic test available, affordable, accurate, and precise in our setting?
- Can we generate a clinical sensible estimate of our patients pre-test probability?
- Will the resulting post-test probabilities affect our management and help our patient?

Generate a clinical sensible estimate of our patient's pre-test probability

- From clinical experience, prevalence statistics, practice databases, this report, or other studies designed for pretest Probability.
- Are the study pa
- Is it unlikely that probabilities have was gathered?

TABLE 2. Pretest Likelihood of CAD in Symptomatic Patients According to Age and Sex*

	Nonanginal Chest Pain		Atypical Angina		Typical Angina	
Age, y	Men	Women	Men	Women	Men	Women
30–39	4	2	34	12	76	26
40–49	13	3	51	22	87	55
50–59	20	7	65	31	93	73
60–69	27	14	72	51	94	86

*Each value represents percent with significant CAD on catheterization.

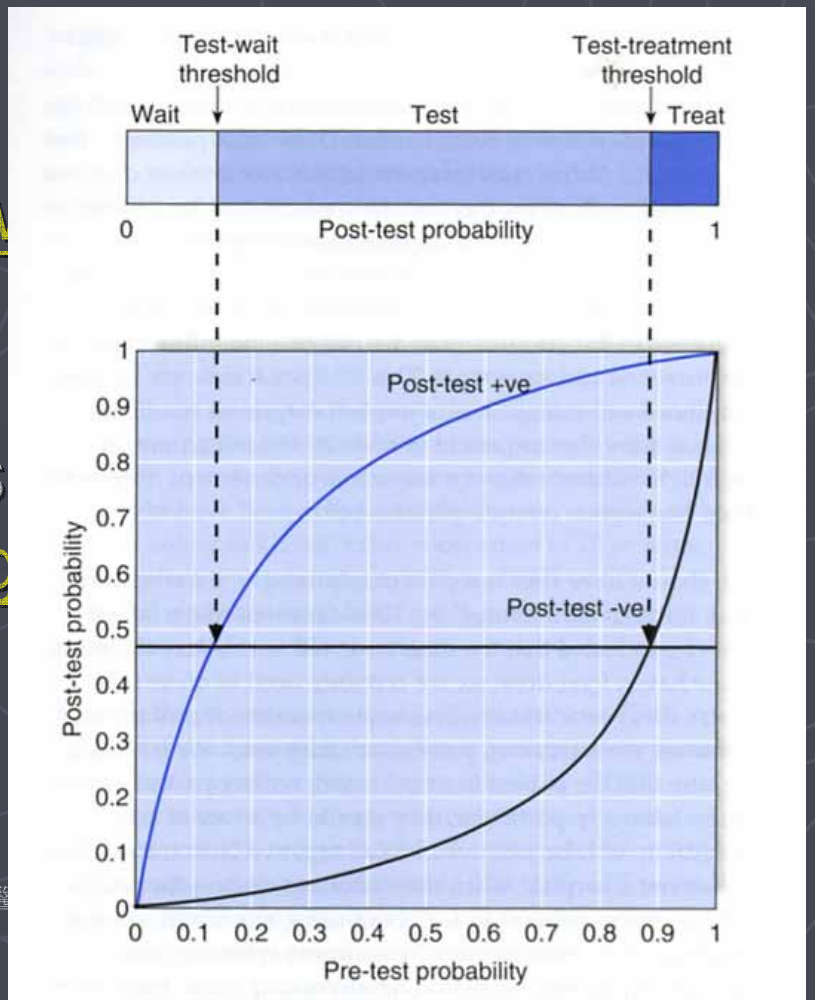
Data from (1) Diamond GA, Forester JS. Analysis of probability as an aid in the clinical diagnosis of coronary-artery disease. *N Engl J Med.* 1979;300:1350–1358. (2) Chaitman BR, Bourassa MG, Davis K, Rogers WJ, Tyras DH, Berger R, Kennedy JW, Fisher L, Judkins MP, Mock MB, Killip T. Angiographic prevalence of high-risk coronary artery disease in patient subsets (CASS). *Circulation.* 1981;64:360–367.

Generate a clinical sensible estimate of our patient's pre-test probability

- From clinical experience, prevalence statistics, practice databases, this report, or other studies designed for pretest Probability.
- Are the study patients similar to our own?
- Is it unlikely that the disease possibilities or probabilities have changed since this evidence was gathered?

Will the Resulting Post-test Probabilities Affect Our Management and Help Our Patient?

- Could it move us across a test-treatment threshold?
- Would our patient be a w
- Would the consequences
  patient reach his or her g



Will the Resulting Post-test Probabilities Affect Our Management and Help Our Patient?

- Could it move us across a test-treatment threshold?
- Would our patient be a willing partner in test?
- Would the consequences of the test help our patient reach his or her goals of therapy?

Therapy

Critically Appraising Treatment article (VIP)

- Validity
  - Is it valid? (closeness to the truth)
- Important
  - Is it important? (size of effect)
- APply
  - Is it applicable to the (specific) patient? (clinical applicability)

Types of Study

- Randomized controlled Trials
- Cohort Studies
- Case Control Studies
- Case reports and case series
- Systematic reviews
  - Meta-analysis: combining many studies into one
- ……

證據的分級


N of 1 RCT
全盤性文獻回顧 Systematic review
綜合分析Meta-analysis
前瞻式隨機分派控制型試驗 RCT
前瞻式非隨機分派控制型試驗
前瞻式世代型研究 Cohort
病例控制世代型研究 Case control
橫斷式調查分析 Cross-sectional
病例(系列)報告Case reports
Unsystemic observation of individual clinician

楔子 – 問題思考

例 1：王老先生與 aspirin－

王老先生罹患高血壓服藥 10 年。
近 5 年來，醫師加上 **aspirin100mg 1#qd**。
上個月，王老先生因 **腦出血** 送醫院急救無效，二天後 **逝世**。

事後省思，要是醫師後來**沒有**加上 **aspirin**，就**不會**發生 **腦出血** 事件？

例 2：多吃鈣片可以長高…

小明 **10 歲**的時候，有 **150cm** 高，由於媽媽**每天**叫他**吃鈣片**，
18 歲時長到 **185cm**。所以，多吃鈣片可以長高？

此兩個案例對 **因果關係** 的 **推論**，有何**問題**？
在**科學**上(醫學上)，如何 証實 **A 事件** 與 **B 事件** 的 **因果關係**？

答案： － 最少要有 **對照個案** 的 **比較**

對照個案 的 比較

例 1：王老先生與李老先生…

王老先生	同鄉，住隔壁… 相同年齡、相同體態、 相同生活習慣與環境	有使用 aspirin	發生 **腦出血**
李老先生		沒有使用 aspirin	未發生 **腦出血**

是否可以用 aspirin 藥物之使用與否，來解釋腦出血的事件？

例 2：多吃鈣片，可以長高？

小明	住隔壁兩個 10 歲小孩: 相同年齡、讀同一班、 相同生活環境、一樣高	每天吃鈣片	**18 歲時 185cm**
小華		從來不吃鈣片	**18 歲時 162cm**

對照個案 比較 的 **兩個問題：**

1. 在現實 (生物) 世界中，充滿許多未知的影響因素，

   導致**隨機變異 (random variation)** (或**隨機誤差**) 處處可見，例：

   1) 同樣每天吃鈣片，長成的身高，有高有矮；

   2) 再怎樣細心，同一管血在不同時候測得的血糖，多少還是有變異。

在 醫學研究中，如何 **克服 隨機變異** 的 **問題**？

答：增加 **樣本數 (重覆測量)**，不以 **單一個案** 來 **比較**，

而是以 **樣本平均值 (點估計值)** 來 **比較**，以 **除去** 隨機誤差。

2. 兩組個案，不見得可以完全相比擬…。

   可比較性(comparability)不足 = 存在**干擾因素** (confounding factors)

   何謂 **干擾因素**？

   實驗組 與 **對照組** 之間，有某些 **不同** 之 **因素**，

   這些因素 **與結果** 變項(疾病) **相關**，而可解釋「**所觀察的現象**」例如

1. 調查社區健康情形：**運動量**多的居民，**心臟病**的比例高。Why？

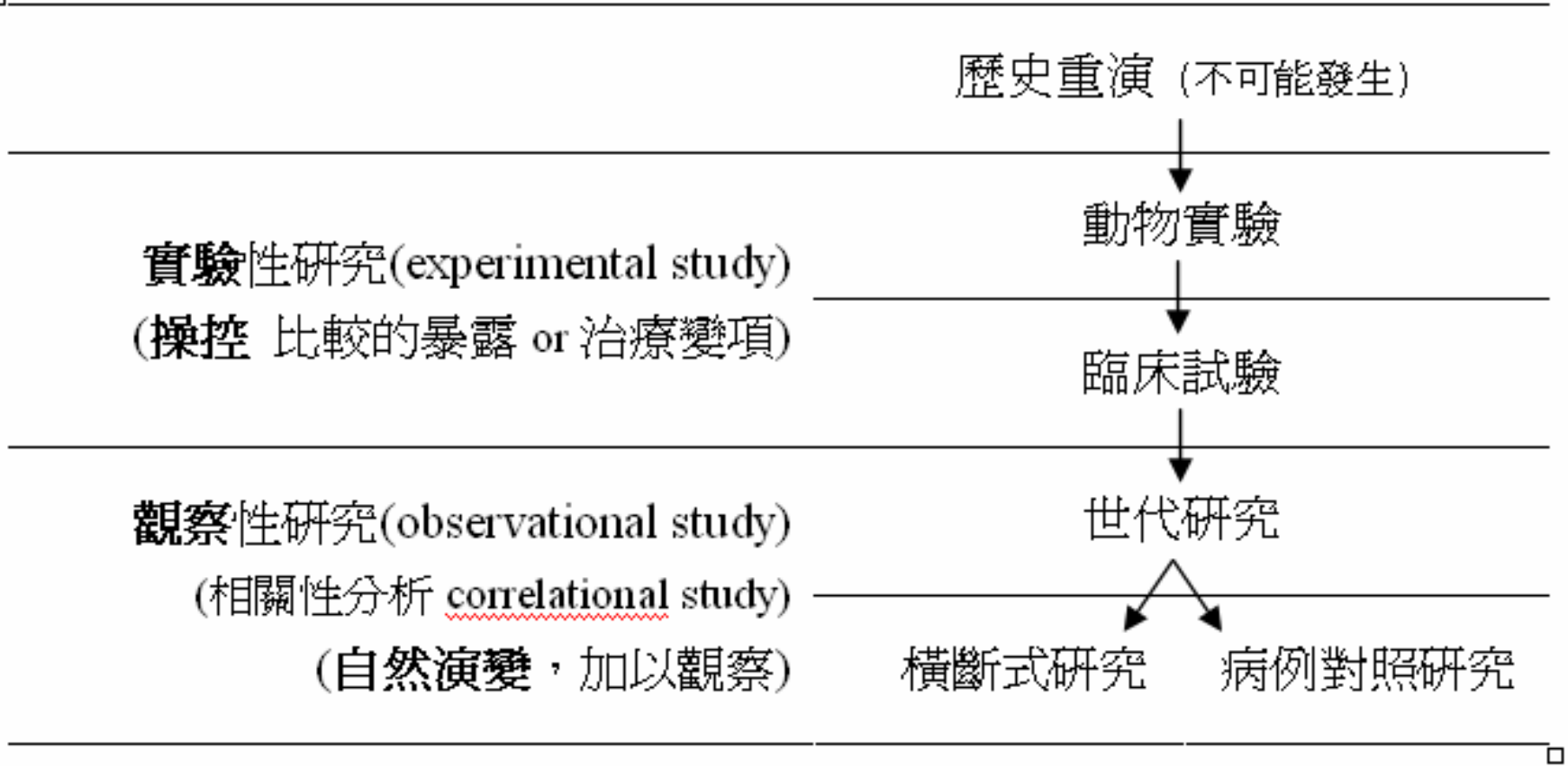
對照組的選取：著眼在「可比較性」

歷史重演 (Counter-factual)

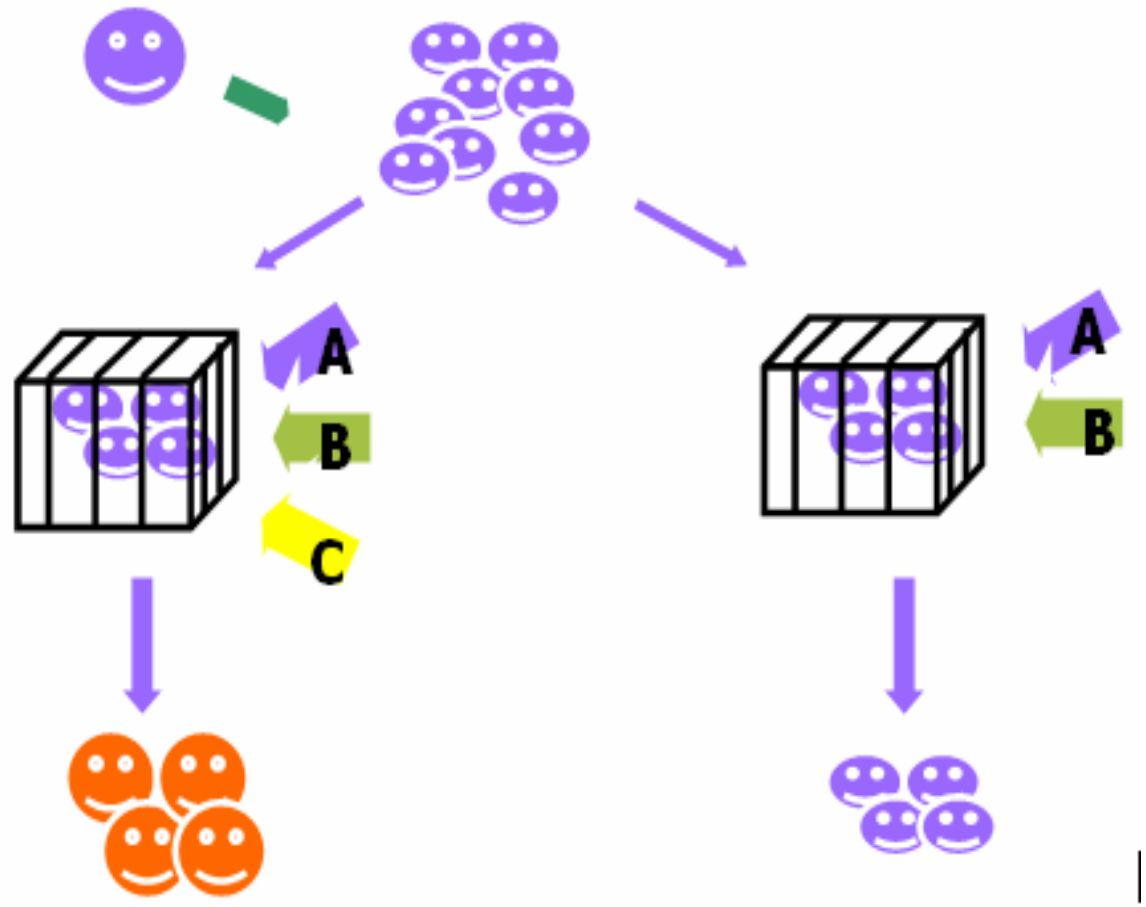
解決「**可比較性**」的問題，最好的辦法是「**歷史重演－自己和自己比**」，但是，**違逆事實，不可能發生**…

由於歷史無法重演，只好用其他的方法，獲得還算可信 (可比較) 的對照組：

實驗操控 vs. 自然變異



動物實驗 (animal experiment)



操控實驗有興趣的變項，有三個特點：

1. **對照組**的**比較**(Control group comparison)，盡可能維持**兩組相同**的**處置**。

2. **隨機分派** Randomization process 的過程

   隨機 將 受試個體 **分派** 至 **實驗組/對照組**，使 **可能有而未知** 之 **個別差異**，盡可能平均分配，使實驗具「**可比較性**」**(減少干擾因素)**；

3. 盲目程序 Blinded procedure

   對**結果**進行**測量者**，不知道**隨機分派**中，**受試者**所分到的**組別**。
   避免 在**測量**結果時，受到 **主觀成見** 的 **影響**。

臨床試驗 (randomized double blinded clinical trial)

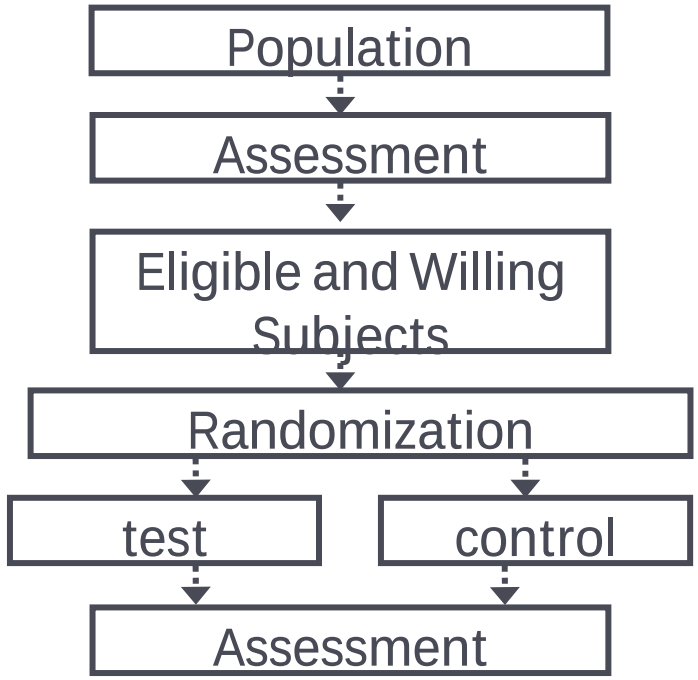
在**動物實驗**中，我們已經**盡可能**地**控制**實驗**前後**的**影響** (干擾) **因素**，
所以可能影響結果的因素不多。因此，實驗樣本的數目不必太多。

但在有關人的**臨床試驗**中，有很多無法操控的變項(干擾因素)，包括：
受試個體特質 / **環境** / **時間**等。

例如：無法強迫一家人都參與實驗、不能把人關在籠子、控制食物的量與種類、控制每天的運動量…

只得 **增加** 實驗 **受試者** 的 **數目**，並 **仔細** 地 **隨機分派**，
使得 **已知** / **未知**的 干**擾因素**，盡可能地**平均分配**，**減少干擾**可能。

如此，**實驗組** 與 **對照組** 之間，得以具備「**可比較性**」。



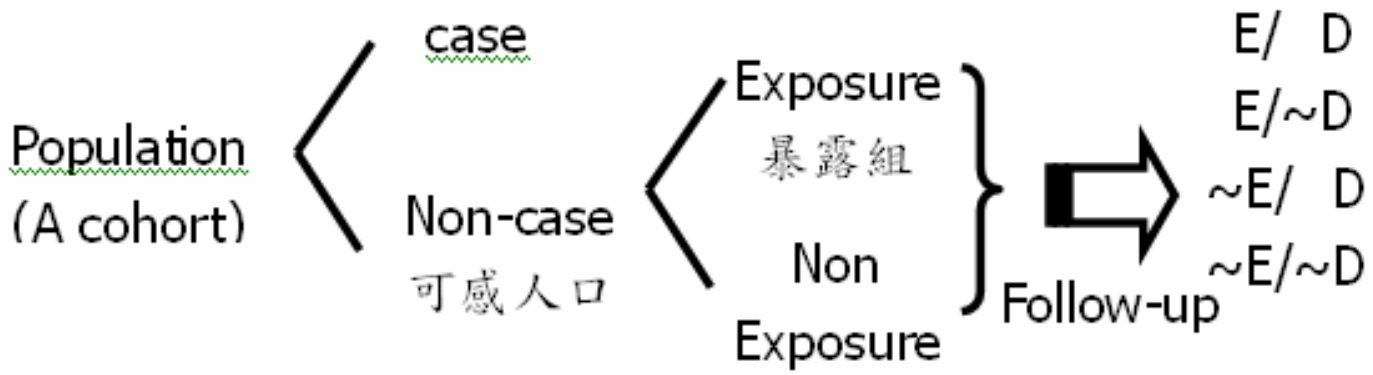
結果（二元變項：有病/沒病）呈現如下：

	有病	沒病	發病危險性
實驗組 (Exposure) / n=500	15	485	0.03 = 3%
對照組 (Non-Exposure) / n=500	5	495	0.01 = 1%
Relative risk (R.R)	0.03/0.01		= 3.0

困難：花錢、花時間，在日常生活中，病患很難配合，也緩不濟急…

觀察性研究 (observational study)

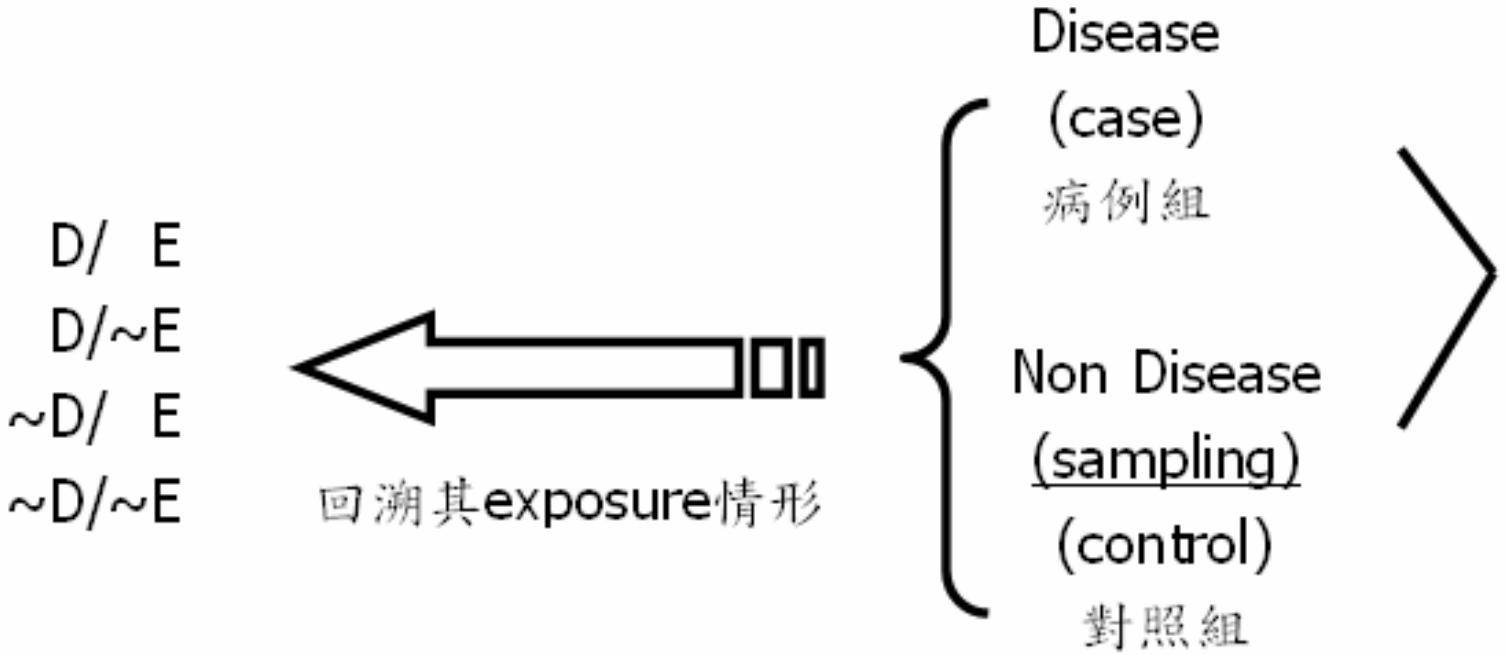
1. 世代追蹤研究 (Cohort study)



例如：經調查發現，某社區 50~59 歲女性族群共 3978 人，其中 1789 人服用停經女性荷爾蒙，另 2189 人沒人服用。追蹤 10 年後，發生乳癌之病例，前者有 15 人，後者有 5 人。發生乳癌的比率，似乎比較高。

Cohort population 3978 人	有病	沒病	發病比率
Exposure 1789 人	15	1774	= 0.84%
Non-Exposure 2189 人	5	2184	= 0.23%

2. 病例對照研究 (Case control study)



	E	~E
D	a	b
~D	c	d

關心：勝算比 **odds ratio(OR) = (a/b) / (c/d) = ad/bc**

(危險對比值)

例如：

假設某醫院**新生嬰兒**流行**某疾病**，懷疑與**母親**懷孕時**飲食/用藥**有關…

我們收集一整年，發現有 **20 個有病**的**新生嬰兒**…

	有病	全族群	**沒病**	全族群	發病比率	勝算 =
	20 人	N	**20 人**	N	= 20 / N	有病/沒病
Exposure	**15**	N_E	**8 人**	N_E =**8X**	= **15** / **8X**	1.875
Non-Exposure	**5**	N_{NE}	**12 人**	N_{NE}=**12X**	= **5** / **12X**	0.417
Relative risk	(R.R)	$= \frac{15 / 8X}{5 / 12X} = \frac{15/8}{5/12} = 4.50$				勝算比 = 4.5

用在罕見的疾病、快速地找出可能的病因 (網漁流行病學)

假設比較多，因果關係不能肯定。多需要更進一步研究証實。

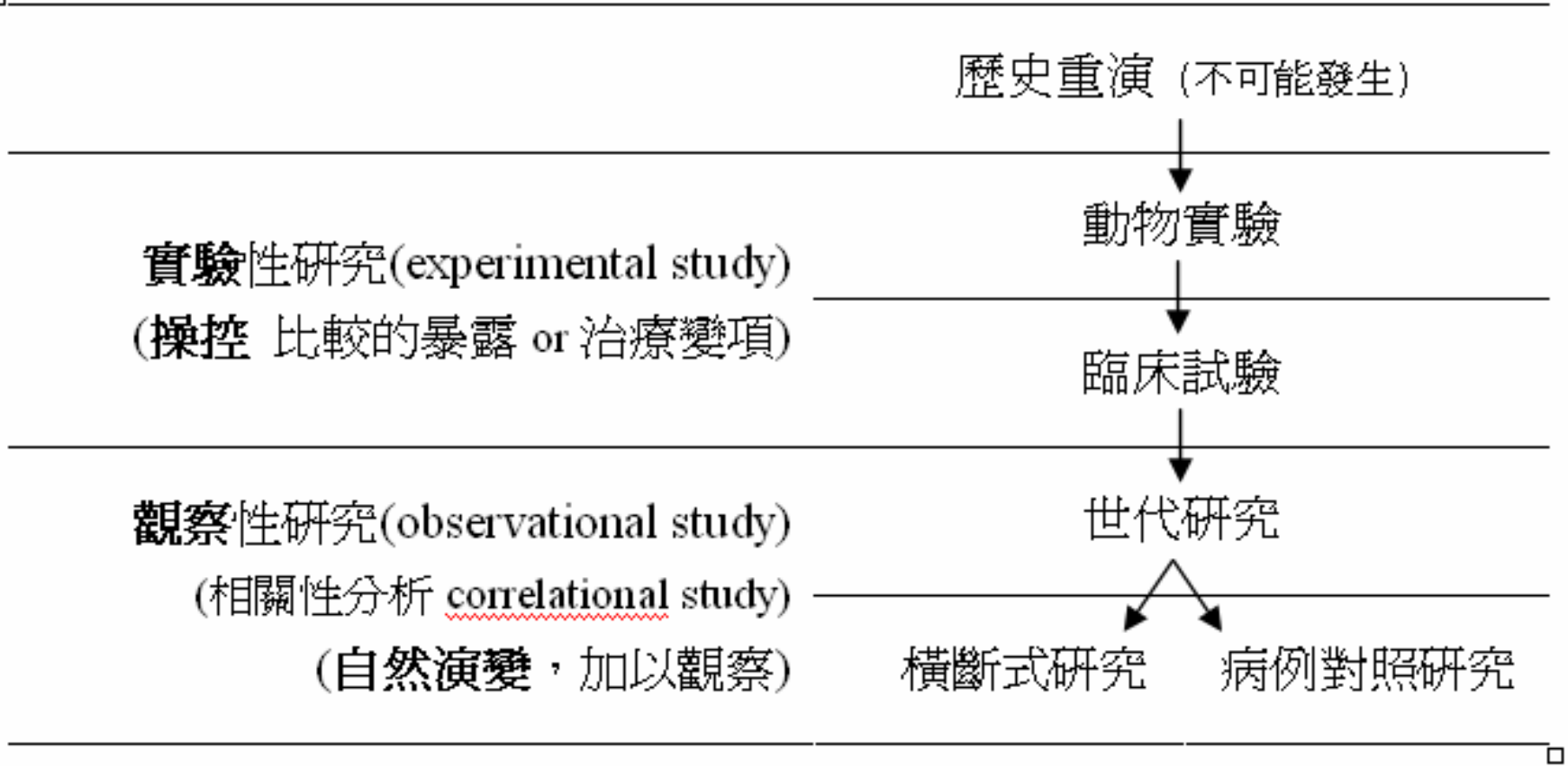
對照組的選取：著眼在「可比較性」

歷史重演 (Counter-factual)

解決「**可比較性**」的問題，最好的辦法是「**歷史重演－自己和自己比**」，但是，**違逆事實，不可能發生**…

由於歷史無法重演，只好用其他的方法，獲得還算可信 (可比較) 的對照組：

實驗操控 vs. 自然變異



配對與交叉設計

思考：在某些情況下，我們可以讓歷史重演

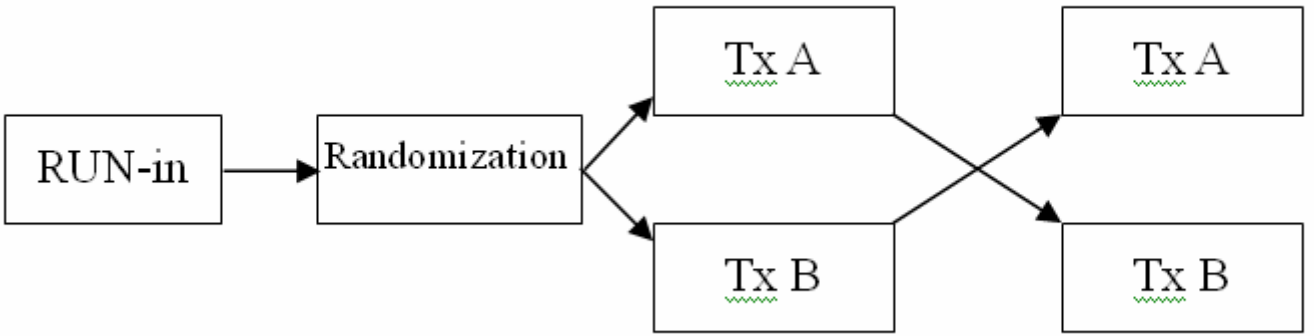
(自己如何可能和自己比較？)

洗髮精 的 廣告

PST 試驗

腎高血壓 與 腎血管病變

降血壓藥物 與 血壓 之 臨床試驗



Is this evidence about therapy valid?

(from Double blinded Randomized controlled Clinical Trial)

- Randomization of assignment of pts to different tx
- Blinded procedure:
  - Concealed the randomization in enrollment
  - Kept the pts, Drs, and study personnel blinded to tx grp
- Compatibility of treatment vs. control groups:
  - Were the groups similar at the start of the trial?
  - Were groups treated equally, apart from the experimental therapy?
  - Were all patients analyzed in the groups to which they were randomized? *(intention-to-treat analysis)*
- Sufficiently long and complete follow-up?

Is this valid evidence about tx important?

- What is the magnitude of the treatment effect?
- How precise is the estimate of the treatment effect?

組別(Group) / 有病人數	總人數	發病危險性
實驗組 (Treatment)/ Ed	TotalE	Experimental Event Rate (EER) = Ed / TotalE
對照組 (Control) / Cd	TotalC	Control Event Rate (CER) = Cd / TotalC
Relative Risk Reduction(RRR) =		(CER-EER) / CER
Absolute risk reduction(ARR) =		(CER-EER)
Number needed to treat (NNT) =		1 / ARR = 1 / (CER-EER)

組別(Group) / 有病人數	總人數	發病危險性	虛擬研究結果
實驗組 (Treatment) / 10	500	10/500 = 0.02 = 2%	10/5000 = 0.002 = 0.2%
對照組 (Control) / 15	500	15/500 = 0.03 = 3%	15/5000 = 0.003 = 0.3%
Relative Risk Reduction(RRR) =		(3%-2%)/3% = 0.33	(0.3%-0.2%)/0.3%=0.33
Absolute risk reduction(ARR) =		(3%-2%) = 1%	(0.3%-0.2%) = 0.1%
Number needed to treat (NNT) =		1/1% = 100	1/0.1% = 1000

Table 5.4 Some useful NNTs[a]

Target disorder	Intervention	Events being prevented	Event rate CER	Event rate EER	Follow-up time	NNT
Diastolic blood pressure 115–129 mmHg[b]	Antihypertensive drugs	Death, stroke, or MI	13%	1.4%	1.5 years	8
Diastolic blood pressure 90–109 mmHg[c]	Antihypertensive drugs	Death, stroke or MI	5.5%	4.7%	5.5 years	128
Symptomatic high-grade carotid stenosis[d]	Carotid endarterectomy (compared with medical therapy)	Death or major stroke	18%	8%	2 years	10
Mild-to-moderate Alzheimer's dementia[e]	Donepezil (vs. placebo)	No functional decline	44%	59%	1 year	7
Unstable angina[f]	Invasive management within 7 days (compared with medical management)	Death or MI	16%	12%	24 months	24
Renal insufficiency and undergoing coronary angiogram[g]	Oral acetylcysteine (vs. placebo)	Contrast media-induced reduction in renal function	12%	4%	48 hours	12

[a]See www.cebm.utoronto.ca for additional NNTs.
[b]JAMA 1967; 202: 116–22.
[c]BMJ 1995; 291: 97–104.
[d]N Engl J Med 1991; 325: 445–53.
[e]Neurology 2001; 57: 613–20.
[f]J Am Coll Cardiol 2002; 40: 1902–14.
[g]JAMA 2003; 289: 553–8.

$$NNT_{hypothetical} \times time_{hypothetical} = NNT_{observed} \times time_{observed}$$

$$NNT_{hypothetical} = NNT_{observed} \times (time_{observed} / time_{hypothetical})$$

$$= 128 \times (5.5 / 1.5) = 470$$

組別(Group) / 有病人數	總人數	發病危險性
實驗組 (Treatment)/ Ed	TotalE	Experimental Event Rate (EER) = Ed / TotalE
對照組 (Control) / Cd	TotalC	Control Event Rate (CER) = Cd / TotalC

Relative risk increased (RRI) = (EER-CER)/CER

Absolute risk increased(ARI) = (EER-CER)

Number needed to Harm(NNH) = 1/ARI = 1/(EER-CER)

組別(Group) / 有病人數	總人數	發病危險性
實驗組 (Treatment) / 15	500	15/500 = 0.03 = 3%
對照組 (Control) / 10	500	10/500 = 0.02 = 2%

Relative risk increased (RRI) = (3%-2%)/2% = 0.5

Absolute risk increased(ARI) = (3%-2%) = 1%

Number needed to Harm(NNH) = 1/1% = 100

Is the evidence applicable to our patient?

- Is our patient so different from those in the study that its results cannot apply?
  - Fit all the inclusion criteria for the study / different sociodemographic features or pathobiology (pharmacogenetics, absent immune responses, …)
- Is the treatment feasible in our setting?
  - available in our setting/ payed by?/ administration & required monitoring…
- What are our patient's potential benefits and harms from the therapy?
  - NNT = 1/(PEER x RRR) or $NNT_{patient} = NNT_{study}/f_t$
  - NNH = 1/(PEER x RRI) or $NNH_{patient} = NNH_{study}/f_h$
- What are our patient's values and expectations for both the outcome we are trying to prevent and the treatment we are offering?

組別(Group) / 有病人數	總人數	發病危險性
實驗組 (Treatment)/ Ed	TotalE	Experimental Event Rate (EER) = Ed / TotalE
對照組 (Control) / Cd	TotalC	Control Event Rate (CER) = Cd / TotalC
Relative Risk Reduction(RRR) =	(CER-EER) / CER	
Absolute risk reduction(ARR) =	(CER-EER)	
Number needed to treat (NNT) =	1 / ARR = 1 / (CER-EER)	

- NNT = 1/ARR = 1/(CER-EER) = 1/(CER x RRR)

組別(Group) / 有病人數	總人數	發病危險性	Your patient's condition
實驗組 (Treatment) / 10	500	10/500 = 0.02 = 2%	
對照組 (Control) / 15	500	15/500 = 0.03 = 3%	PEER= 0.003, f_t=0.1
Relative Risk Reduction(RRR) =		(3%-2%)/3% = 0.33	(0.3%-0.2%)/0.3%=0.33
Absolute risk reduction(ARR) =		(3%-2%) = 1%	(0.3%-0.2%) = 0.1%
Number needed to treat (NNT) =		1/1% = 100	1/(0.003x0.33) = 1000 100 / 0.1 = 1000

- NNT = 1/ARR = 1/(CER-EER) = 1/(CER x RRR)
- NNT = 1/(PEER x RRR)

  PEER : Patient's Expected Event Rate with control tx
- $NNT_{patient} = NNT_{study}/f_t$

  f_t : Risk of the outcome in your patient, relative to pts in trial.

The likelihood of being helped vs. harmed (LHH)

- (1 / NNT) x f_t x S : (1/NNH) x f_h

  = ARR x f_t x S : ARI x f_h

f_t : Risk of the disease in your patient, relative to pts in trial.

f_h : Risk of the side effect in your patient, relative to pts in trial.

S : severity factor, the relative severity of
disease progression if no treatment
to side effect if receiving treatment.

The likelihood of being helped vs. harmed (LHH)

研究文獻	總人數	發病人數	發病危險		副作用人數	副作用危險
實驗組	500	10	2.00%		15	3.00%
對照組	500	15	3.00%		10	2.00%
		PEER =	9.00%		PEER =	0.66%
		ft =	3		fh =	0.33
		S =	2			1
		RRR =	0.33333		RRI =	0.50000
		ARR =	0.01000		ARI =	0.01000
		NNT =	100.000		NNH =	100.000
		LHH =	= ARR x ft x S : ARI x fh =			18.18

MRC/BHF Heart Protection Study of cholesterol lowering with simvastatin in 20 536 high-risk individuals: a randomised placebo-controlled trial

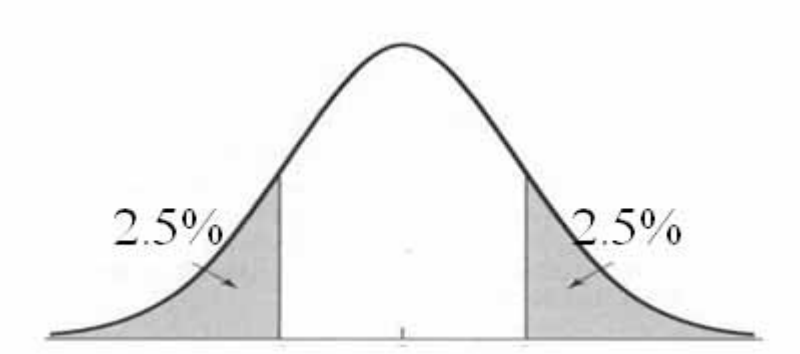
*Heart Protection Study Collaborative Group**

研究文獻	總人數	發病人數	發病危險		副作用人數	副作用危險
實驗組	10269	442	4.30%		5	0.05%
對照組	10267	585	5.70%		3	0.03%
		PEER =	17.09%		PEER =	0.01%
		ft =	3		fh =	0.33
		S =	19			1
		RRR =	0.24459		RRI =	0.66634
		ARR =	0.01394		ARI =	0.00019
		NNT =	71.754		NNH =	5136.001
		LHH =	= ARR x ft x S : ARI x fh =			12363.45

Is this valid evidence about tx important?

- What is the magnitude of the treatment effect?

- How precise is the estimate of the treatment effect? *(95% confidence interval)*



謝謝聆聽！