實證文獻等級介紹

光田綜合醫院 家醫科 王牧羣醫師 2009/3/19



Level of evidence

證據等級

- Confidence of the resources we found
- According to <u>validity</u> of the study
- "RAMbo"
 - 研究族群是否具有代表性(Representative)?隨機選擇(random selection)/隨機分派(random allocation)
 - -是否有足夠的確認與追蹤(Ascertainment)?反應率/追蹤/確認>80%

Grade of recommendation

建議強度

- The strength of recommendation of a <u>clinical bottom line</u> derived from summation of individual evidence
 - Dependent on level of evidence



OXFORD CENTRE FOR EVIDENCE-BASED MEDICINE LEVELS OF EVIDENCE

- -Most commonly applied in EBM
- -http://www.cebm.net/levels_of_evidence.asp



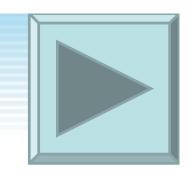


Five Categories

- Therapy/Prevention, Aetiology/Harm
 - Prognosis
- Diagnosis
- Diagnosis
 Differential diagnosis/symptom prevalence study
 Economic and decision analyses



Therapy/Prevention, Aetiology/Harm



1a	SR (with <u>homogeneity*</u>) of RCTs
1b	Individual RCT (with narrow Confidence Interval‡)
1c	All or none§
2a	SR (with <u>homogeneity*</u>) of cohort studies
2b	Individual cohort study (including low quality RCT; e.g., <80% follow-up)
2c	"Outcomes" Research; Ecological studies
3a	SR (with <u>homogeneity*</u>) of case-control studies
3b	Individual Case-Control Study
4	Case-series (and poor quality cohort and case-control studies§§)
5	Expert opinion without explicit critical appraisal, or based on physiology, bench research or "first principles"



Homogeneity

- A systematic review that is free of worrisome variations (heterogeneity) in the directions and degrees of results between individual studies.
- Not all statistically significant heterogeneity need be worrisome, and not all worrisome heterogeneity need be statistically significant.
- Studies displaying worrisome heterogeneity should be tagged with a "-" at the end of level.



Example: Forest plot

Comparison: 03 Treatment versus Placebo

Outcome: 01 Effect of treatment on mortality

Study	Treatment n/N	Control n/N	OR (95%CI Fixe	Weight ed) %	OR (95%Cl Fixed)
Brown 1998	24 / 472	35 / 499		9.6	0.71[0.42,1.21]
Geoffrey 1997	120 / 2850	182 / 2838		51.8	0.64[0.51,0.81]
Mason 1996	56 / 2051	84 / 2030	_ I	24.4	0.65[0.46,0.92]
Peters 2000	5 / 81	4 / 78		1.1	1.22[0.31,4.71]
Scott 1998	31 / 788	46 / 792		13.1	0.66[0.42,1.06]
Total(95%CI)	236 / 6242	351 / 6237	•	100.0	0.66[0.56,0.78]
Test for heterogeneity chi-	square=0.92 df=4 p=0.9	12			
Test for overall effect z=-	4.82 p<0.00001				
			.1 .2 1	5 10	
			Favours treatment	Favours control	

Eyeball test

Cochran Q

Back to text



"Eyeball" test

- Look for overlap of the confidence intervals of the trials with the summary estimate.
 - In the <u>example</u>, note that the dotted line running vertically through the combined odds ratio crosses the horizontal lines of all the individual studies indicating that the studies are homogenous.



Cochran chi-square(Cochran Q)

- If Cochran Q is statistically significant there is definite heterogeneity.
- If Cochran Q is not statistically significant but the ratio of Cochran Q and the degrees of freedom (Q/df) is > 1 there is possible heterogeneity.
- If Cochran Q is not statistically significant and Q/df is < 1 then heterogeneity is very unlikely.



Cochran chi-square(Cochran Q)

- The level of significance for Cochran Q is often set at 0.1 due to the low power of the test to detect heterogeneity.
- In the <u>example</u> Q/df is <1 (0.92/4= 0.23) and the p-value is not significant (0.92) indicating no heterogeneity.





All or none

 Met when <u>all</u> patients died before the Rx became available, but some now survive on it; or when some patients died before the Rx became available, but <u>none</u> now die on it.





Poor quality cohort study

- Failed to clearly define comparison groups
 - Failed to measure exposures and outcomes in the same (preferably blinded), objective way in both exposed and nonexposed individuals
- Failed to identify or appropriately control known confounders
- known confounders
 Failed to carry out a sufficiently long and complete follow-up of patients

Poor quality case-control study

- Failed to clearly define comparison groups
- Failed to measure exposures and outcomes in the same (preferably blinded), outcomes in the same (preferably blinded objective way in both cases and controls

 Failed to identify or appropriately control known confounders objective way in both cases and controls
 - known confounders





Prognosis



<u></u>	1a	SR (with <a href="https://www.new.new.new.new.new.new.new.new.new.</td></tr><tr><td></td><td>1b</td><td>Individual inception cohort study with <math>\geq</math> 80% follow-up; CDR† validated in a single population</td></tr><tr><td>()</td><td>1c</td><td>All or none case-series</td></tr><tr><td>@</td><td>2a</td><td>SR (with <a href=" https:="" td="" www.norm.new.norm.new.new.new.new.new.new.new.new.new.new<="">
er a	2b	Retrospective cohort study or follow-up of untreated control patients in an RCT; Derivation of CDR† or validated on split-sample§§§ only
	2c	"Outcomes" Research
I	3a	
0	3b	
$\frac{0}{0}$	4	Case-series (and poor quality prognostic cohort studies***)
E E	5	Expert opinion without explicit critical appraisal, or based on physiology, bench research or "first principles"

CDR(Clinical Decision Rule)

 Algorithms or scoring systems which lead to a prognostic estimation or a diagnostic category





Validation(驗證)

Index	G	ibson and Ste	phenson [33]		Feldman et al. [34]			
_	Retro	ospective	Pro	spective	Retr	ospective	Pros	pective
≥ 14 < 14 Total	1.00 0.37 0.47	(29/29) (58/158) (87/187)	1.00 0.28 0.46	(13/13) (11/39) (24/52)	0.70 0.17 0.19	(16/23) (76/456) (92/479)	0.60 0.19 0.22	(6/10) (23/121) (29/131)
PSEP (95 per cent CI)		0.63 5 to 0.71)		0.72 to 0.86)		0.54 1 to 0.72)	0.41 (0.10 to 0.72)	
		ingle da	ata s	et				
• Interna	al: si	O		_	rom t	he san	റല	
	al: si oral:	O		_	rom t	he san	ne	
InternationTempo centerExternation	al: si oral: (s) nal: f	O	d dat	a set f				

- Temporal: second data set from the same



Split-sample validation

 Collecting all the information in a single tranche, then artificially dividing this into "derivation" and "validation" samples





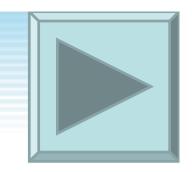
Poor quality prognostic cohort study

- Sampling was biased in favour of patients who already had the target outcome
- The measurement of outcomes was accomplished in <80% of study patients
- Outcomes were determined in an unblinded, non-objective way
- No correction for confounding factors





Diagnosis



1a	SR (with <u>homogeneity*</u>) of Level 1 diagnostic studies; <u>CDR†</u> with 1b studies from different clinical centres
1b	<u>Validating**</u> cohort study with <u>good†††</u> reference standards; or <u>CDR†</u> tested within one clinical centre
1c	Absolute SpPins and SnNouts††
2a	SR (with <a first="" href="https://www.new.new.new.new.new.new.new.new.new.</td></tr><tr><td>2b</td><td>Exploratory** cohort study with good†††reference standards; CDR† after derivation, or validated only on split-sample§§§ or databases</td></tr><tr><td>2c</td><td></td></tr><tr><td>3a</td><td>SR (with homogeneity*) of 3b and better studies</td></tr><tr><td>3b</td><td>Non-consecutive study; or without consistently applied reference standards</td></tr><tr><td>4</td><td>Case-control study, poor or non-independent reference standard</td></tr><tr><td>5</td><td>Expert opinion without explicit critical appraisal, or based on physiology, bench research or " principles"<="" td="">

Validating vs Exploratory

- Validating studies test the quality of a specific diagnostic test, based on prior evidence.
- evidence.

 Exploratory studies collect information and trawls the data (e.g. using a regression analysis) to find which factors are 'significant'.





Reference standards

- Good reference standards are independent of the test, and applied blindly or objectively to applied to all patients.
- Poor reference standards are haphazardly applied, but still independent of the test. Use of a non-independent reference standard (where the 'test' is included in the 'reference', or where the 'testing' affects the 'reference') implies a level 4 study.





SpPin & SnNout

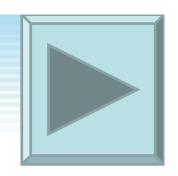
- An "Absolute SpPin" is a diagnostic finding whose <u>Specificity</u> is so high that a <u>Positive</u> result rules-in the diagnosis.
- result rules-in the diagnosis.

 An "Absolute SnNout" is a diagnostic finding whose Sensitivity is so high that a Negative result rules-out the diagnosis.





Differential diagnosis/ symptom prevalence study



1a	SR (with homogeneity*) of prospective cohort studies
1b	Prospective cohort study with good follow-up****
1c	All or none case-series
2a	SR (with homogeneity*) of 2b and better studies
2b	Retrospective cohort study, or poor follow-up
2c	Ecological studies
3a	SR (with homogeneity*) of 3b and better studies
3b	Non-consecutive cohort study, or very limited population
4	Case-series or superseded reference standards
5	Expert opinion without explicit critical appraisal, or based on physiology, bench research or "first principles"



Good follow-up

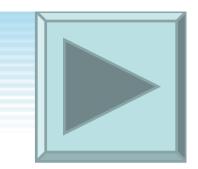
- Follow-up in a differential diagnosis study >80%
 - Adequate time for alternative diagnoses to emerge (eg 1-6 months acute, 1-5 years chronic)





Economic and decision analyses

theory or "first principles"



1a	SR (with homogeneity*) of Level 1 economic studies
1b	Analysis based on clinically sensible costs or alternatives; systematic review(s) of the evidence; and including multi-way sensitivity analyses
1c	Absolute better-value or worse-value analyses ††††
2a	SR (with <a href="https://www.news.news.news.news.news.news.news.n</td></tr><tr><td>2b</td><td>Analysis based on clinically sensible costs or alternatives; limited review(s) of the evidence, or single studies; and including multi-way sensitivity analyses</td></tr><tr><td>2c</td><td>Audit or outcomes research</td></tr><tr><td>3a</td><td>SR (with homogeneity *) of 3b and better studies
3b	Analysis based on limited alternatives or costs, poor quality estimates of data, but including sensitivity analyses incorporating clinically sensible variations
4	Analysis with no sensitivity analysis
5	Expert opinion without explicit critical appraisal, or based on economic

Better-value vs Worse-value

- Better-value treatments are clearly as good but cheaper, or better at the same or reduced cost.
- reduced cost.
 Worse-value treatments are as good and more expensive, or worse and the equally or more expensive.





Addendum

- Users can add a minus-sign "-" to denote the level of that fails to provide a conclusive answer because of:
 - EITHER a single result with a wide Confidence Interval (such that, for example, an ARR in an RCT is not statistically significant but whose confidence intervals fail to exclude clinically important benefit or harm)
 - OR a Systematic Review with troublesome (and statistically significant) heterogeneity.
 - Such evidence is inconclusive, and therefore can only generate Grade D recommendations

不同領域的文獻證據力分級

證據力等級 治療, 病因, 預防 預後 診斷 鑑別診斷, 症 狀盛行率研究 策分析, 決 策分析 Level 1 RCT¹ 的系統性 回顧; 或 Confidence Interval 窄的RCT Interval 窄的RCT Enterval 窄的RCT Enterval 窄的RCT Enterval 窄的RCT Enterval 窄的RCT Enterval 窄的RCT Enterval 窄的原体性 医囊炎 医脑腔的 Enterval 定式 医脑腔的 Enterval 定式 医脑腔的 Enterval 定式 医脓性回顧 Enterval 定式 医脓性回额 完成 医脓性回额 完成 医脓性回额 完成 医脓性回额 完成 医脓性回额 完成 医脓性回额 完成 医脓性回动 完成 医脓性回颌 Enterval 定式 医脓性回颌 是要 医脓 方法或成本的 平式 或性 医心腔						
Level 3 有對照組 (controlled study) 条統性回顧;或 違列 80% 比例的世代研究;或 或以公認標準驗證的世代研究;或 或 驗證的世代研究;或 或 臨床指引。 上evel 1 文獻; 或以公認標準驗證的世代研究;或 臨床指引。 之系統性回顧;或 也較好壞方向的研究。 Level 2 世代研究 的系統性回;或 低品質的 RCT 或追蹤小於 80%;或預後研究。 回溯性世代研究的對照組;或由小族群推測或 驗證的臨床指引;或預後研究。 系統性回顧 上evel 2 文獻; 或追蹤不全之 無	證據力等級	治療, 病因, 預防	預後	診斷		
Confidence Interval 室的RCT 達到 80% 比例的世代研究;或人經驗證的世代研究;或處庭床指引 或以公認標準驗證的世代研究;或庭職者引 顧;或追蹤完整之前瞻世代研究的所究 或比較好壞方向的研究 Level 2 世代研究的系統性回溯或。 (Kan質的 RCT或追蹤小於 80%;或預後研究的 (Kan 質) 以表預後研究的 (Kan 可) 以表預後可) 以表預能的 (Kan 可) 以表有能的 (Kan 可) 以表示有的 (Kan 可) 以表示的 (Kan Tan Tan Tan Tan Tan Tan Tan Tan Tan T	Level 1	RCT¹ 的系統性	世代研究 ² 的	系統性回顧	前瞻世代研究	系統性回顧
Level 2 世代研究 的系統性回;或低品質的 RCT或追蹤小於 80%;或預後研究的對照組;或由 73% 反於群推測或 验證的臨床 31;或預後研究 4 系統性回顧 20 濟世代研究 20 新瞻世代研究 20 新能性回顧 20 新世代研究 20 新世代研究 20 新性四顧 20 新性四顧 20 新世代研究 20 新世代研究 20 新世代研究 30 新世代研究 30 新世代研究 30 大族群推测或 验證的臨床指引;或預後研究 4 Level 3 有對照組 (controlled study) 系統性回顧 20 新世代研究 20 新世代研究 30 大家或本的單究;或生態 (ecological)研究 30 大連續或 40 新究 40 世代研究 40 大法或成本的研究,包括敏感度		回顧;或	系統性回顧;或	Level 1 文獻;	之系統性回	Level 1 證據;
Level 2 世代研究 的系統性回溯性世代研究 或追蹤		Confidence	達到 80% 比	或以公認標準	顧;或追蹤完整	或比較好壞方
Level 2 世代研究 的系統性目標		Interval 窄的RCT	例的世代研	驗證的世代研	之前瞻世代研	向的研究
Level 2 世代研究 的系統性回言或性回言或低品質的 RCT或追蹤小於 或強蹤小於 或的學院組;或由 外族群推測或 验證的臨床指引;或預後研究 4 系統性回顧 Level 2 文獻; 或僅在小族群 驗證的臨床指引;或有後研究 2 上evel 2 文獻; 或重要臨床方法或成本的單一研究;或主態 (ecological)研究 3 Level 3 大蘇性回顧 (controlled study) 系統性回顧 (ecological)研究 3 上evel 3 文獻; 或本的研究 4			究;或	究;或臨床指引	究	
Level 2 世代研究的系統性回溯性代研究性回言或低品質的RCT或追蹤小於 80%;或預後研究的 % 回溯性代研究之系統性回溯之系统性回溯的是不全或重要臨床方数。 上evel 2 文獻;或僅在小族群 60% 一种完成,或有後研究的 60% 上evel 2 文獻;或僅在小族群 60% 上evel 2 文獻;或僅在小族群 60% 上evel 2 文獻;或是鄉 60% 上evel 2 文獻;或成本的單文或有後,可究 上evel 2 文獻;或是鄉 60% 一研究;或有後,研究 一研究;或有後,研究 上evel 3 文獻;或成本的研究,包括敏感度 其他臨床方法或成本的研究,包括敏感度 其他臨床方法或成本的研究,包括数域成本的成本的成本的成本的成本的成本的成本的成本的成本的成本的成本的成本的成本的成			經驗證的臨床			
性回;或低品質的 RCT或追蹤 RCT 中未治療的對照組;或由的數照組;或由小族群推測或驗證的臨床指引;或預後研究 4			指引3			
低品質的 RCT 或追蹤小於 80%;或預後研究 80%;或預後研究 4 如	Level 2	世代研究 的系統	回溯性世代研	系統性回顧	回溯世代研究	系統性回顧
或追蹤小於 80%;或預後研究 % 的對照組;或由 小族群推測或 驗證的臨床指 引;或預後研究 4 「A 「A 「A 「A 「A 「A 「A 「A 「A		性回;或	究;或追蹤	Level 2 文獻;	之系統性回	Level 2 文獻;
N		低品質的 RCT	RCT 中未治療	或僅在小族群	顧;或追蹤不全	或重要臨床方
第% 驗證的 臨床指引;或預後研究 4 (ecological)研究 研究 Level 3 有對照組 (controlled study) 系統性回顧 上evel 3 文獻;或不連續或缺 不連續或小族群的世代研究 或成本的研究,包括敏感度		或追蹤小於	的對照組;或由	驗證的臨床指	之回溯世代研	法或成本的單
引;或預後研究 究 Level 3 有對照組 (controlled study) 系統性回顧 Level 3 文獻; 或不連續或缺 不連續或小族 群的世代研究 或成本的研究 ,包括敏感度		80% ; 或預後研	小族群推測或	31	究;或生態	一研究;或預後
Level 3 有對照組 系統性回顧 不連續或小族 其他臨床方法 (controlled study) Level 3 文獻; 或不連續或缺 群的世代研究 ,包括敏感度		究%	驗證的 臨床指		(ecological)研	研究
(controlled study) Level 3 文獻; 群的世代研究 或成本的研究 或不連續或缺 ,包括敏感度			引;或預後研究		究	
(controlled study) Level 3 文獻; 群的世代研究 或成本的研究 或不連續或缺 ,包括敏感度			4			
或不連續或缺 ,包括敏感度	Level 3	有對照組		系統性回顧	不連續或小族	其他臨床方法
		(controlled study)		Level 3 文獻;	群的世代研究	或成本的研究
4 3 4 3 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4				或不連續或缺		,包括敏感度
				乏公認標準驗		(sensitivity) 分
證的研究 析				證的研究		析
Level 4 病例系列 病例系列 對照病例研究 病例系列 未分析敏感度	Level 4	病例系列	病例系列	對照病例研究	病例系列	未分析敏感度
(case- control				(case- control		
study)						
Level 5 專家意見 專家意見 專家意見 專家意見 專家意見 專家意見						

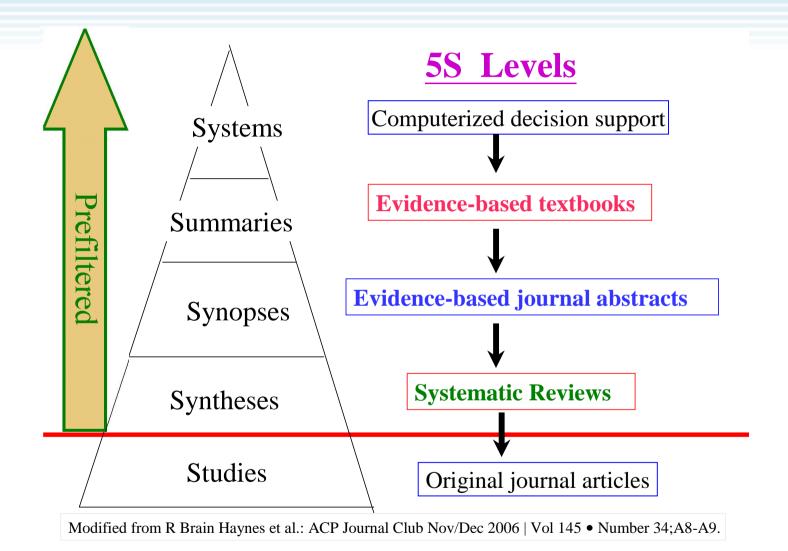
國泰醫院劉致和醫師 2005 年根據牛津實證醫學中心 (Oxford Center for EBM, May 2001) 的列表摘譯 (網址為 http://www.cebm.net/levels of evidence.asp; 原出於 NHS R&D 圓隊 Bob Philips and Chris Ball et al, since 1998)



Grades of Recommendation

Α	consistent level 1 studies
В	consistent level 2 or 3 studies <i>or</i> extrapolations from level 1 studies
С	level 4 studies <i>or</i> extrapolations from level 2 or 3 studies
D	level 5 evidence <i>or</i> troublingly inconsistent or inconclusive studies of any level

 "Extrapolations" are where data is used in a situation which has potentially clinically important differences than the original study situation





THANK YOU FOR YOUR ATTENTION.

