Kuang Tien General Hospital

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

光田綜合醫院 Kuang Tien General Hospital

Level of evidence
證據等級

- Confidence of the resources we found
- According to validity of the study
- “RAMbo”
  - 研究族群是否具有代表性(**R**epresentative)？隨機選擇(random selection)/隨機分派(random allocation)
  - 是否有足夠的確認與追蹤(**A**scertainment)？反應率/追蹤/確認>80%
  - 結果的估計值(**M**easurement)是否公正？盲法(**b**linded)或客觀的(**o**bjective)估計

Grade of recommendation
建議強度

- The strength of recommendation of a clinical bottom line 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 Link

Five Categories

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

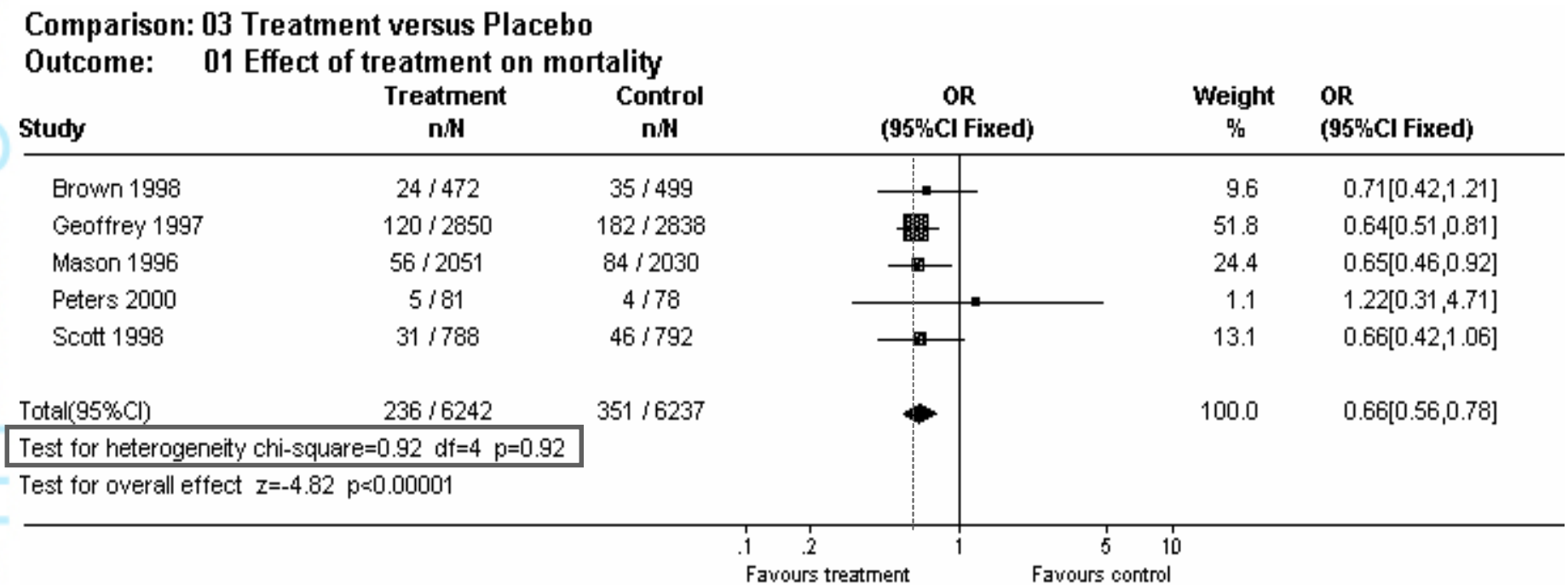
Therapy/Prevention, Aetiology/Harm

1a	SR (with homogeneity*) of RCTs
1b	Individual RCT (with narrow Confidence Interval‡)
1c	All or none§
2a	SR (with homogeneity*) of cohort studies
2b	Individual cohort study (including low quality RCT; e.g., <80% follow-up)
2c	"Outcomes" Research; Ecological studies
3a	SR (with homogeneity*) 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



Eyeball test

Cochran Q

Back to text

"Eyeball" test

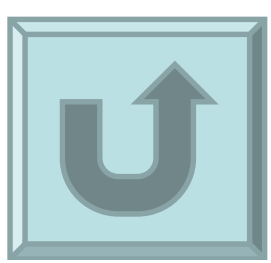
- Look for overlap of the confidence intervals of the trials with the summary estimate.
- In the example, note that the dotted line running vertically through the combined odds ratio crosses the horizontal lines of all the individual studies indicating that the studies are homogenous.

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.

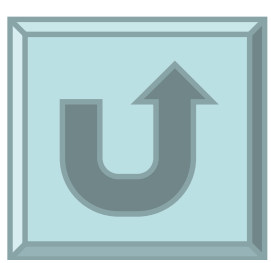
Cochran chi-square(Cochran Q)

- 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 example 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 *all patients died before the Rx* became available, but some now survive on it; or when some patients died before the Rx became available, but *none now die on it*.

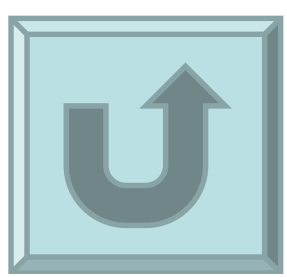


Poor quality cohort study

- Failed to clearly define comparison groups
- Failed to measure exposures and outcomes in the same (preferably blinded), objective way in both exposed and non-exposed individuals
- Failed to identify or appropriately control 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), objective way in both cases and controls
- Failed to identify or appropriately control known confounders

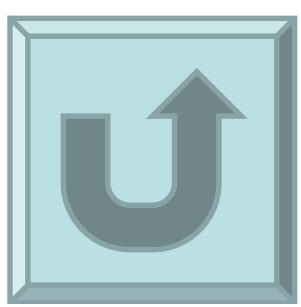


Prognosis

1a	SR (with homogeneity*) of inception cohort studies; CDR† validated in different populations
1b	Individual inception cohort study with ≥ 80% follow-up; CDR† validated in a single population
1c	All or none case-series
2a	SR (with homogeneity*) of either retrospective cohort studies or untreated control groups in RCTs
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
3a	
3b	
4	Case-series (and poor quality prognostic cohort studies***)
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



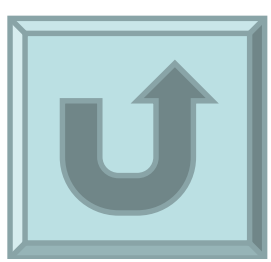
Validation(驗證)

Index	Gibson and Stephenson [33]				Feldman *et al.* [34]			
	Retrospective		Prospective		Retrospective		Prospective	
⩾ 14	1.00	(29/29)	1.00	(13/13)	0.70	(16/23)	0.60	(6/10)
< 14	0.37	(58/158)	0.28	(11/39)	0.17	(76/456)	0.19	(23/121)
Total	0.47	(87/187)	0.46	(24/52)	0.19	(92/479)	0.22	(29/131)
PSEP (95 per cent CI)	0.63 (0.56 to 0.71)		0.72 (0.59 to 0.86)		0.54 (0.34 to 0.72)		0.41 (0.10 to 0.72)	

- Internal: single data set
- Temporal: second data set from the same center(s)
- External: from other centers, perhaps by different investigators

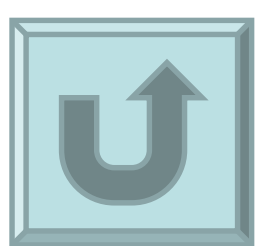
Split-sample validation

- 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

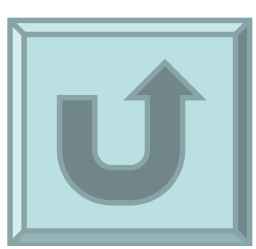


Diagnosis

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

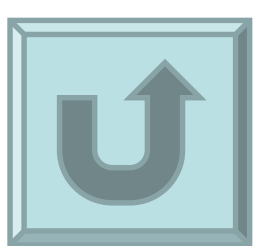
Validating vs Exploratory

- Validating studies test the quality of a specific diagnostic test, based on prior evidence.
- Exploratory studies collect information and trawls the data (e.g. using a regression analysis) to find which factors are 'significant'.



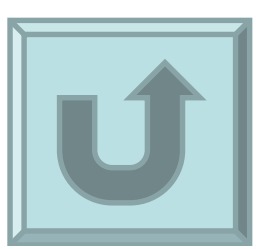
Reference standards

- 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 Specificity is so high that a Positive 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.

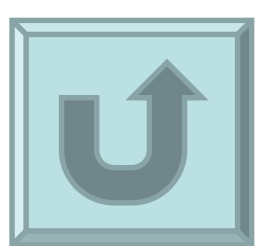


Differential diagnosis/ symptom prevalence study

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)

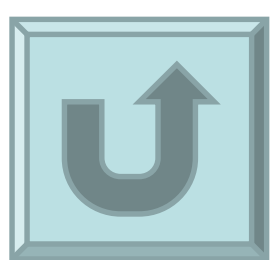


Economic and decision analyses

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 homogeneity*) of Level >2 economic studies
2b	Analysis based on clinically sensible costs or alternatives; limited review(s) of the evidence, or single studies; and including multi-way sensitivity analyses
2c	Audit or outcomes research
3a	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 theory or "first principles"

Better-value vs Worse-value

- Better-value treatments are clearly as good but cheaper, or better at the same or 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

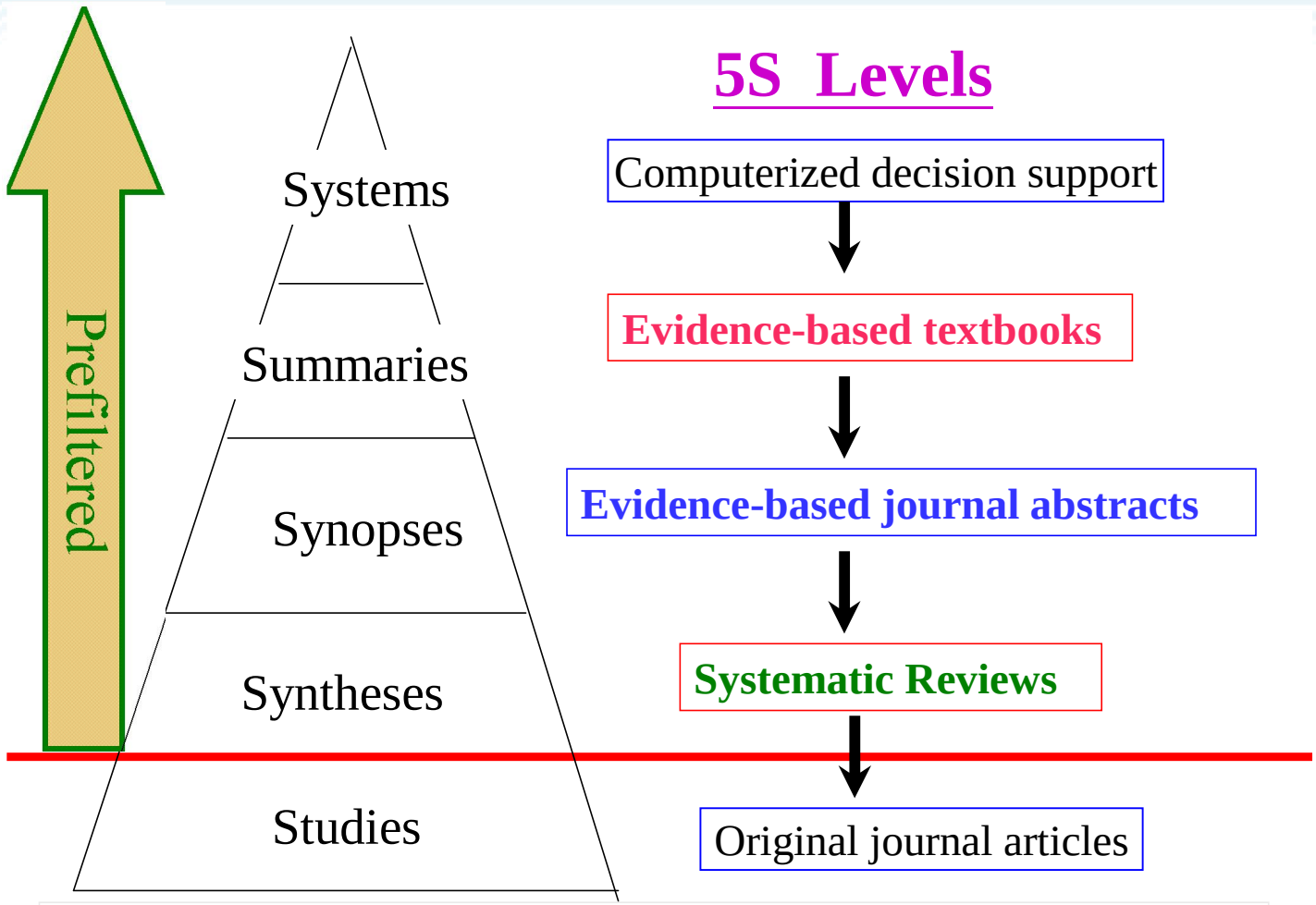
證據力等級	治療，病因，預防	預後	診斷	鑑別診斷，症狀盛行率研究	經濟分析，決策分析
Level 1	RCT[1] 的系統性回顧；或 Confidence Interval 窄的RCT	世代研究[2] 的系統性回顧；或達到 80% 比例的世代研究；或經驗證的臨床指引[3]	系統性回顧 Level 1 文獻；或以公認標準驗證的世代研究；或臨床指引	前瞻世代研究之系統性回顧；或追蹤完整之前瞻世代研究	系統性回顧 Level 1 證據；或比較好壞方向的研究
Level 2	世代研究 的系統性回；或低品質的 RCT 或追蹤小於 80% ；或預後研究%	回溯性世代研究；或追蹤 RCT 中未治療的對照組；或由小族群推測或驗證的 臨床指引；或預後研究[4]	系統性回顧 Level 2 文獻；或僅在小族群驗證的臨床指引	回溯世代研究之系統性回顧；或追蹤不全之回溯世代研究；或生態(ecological)研究	系統性回顧 Level 2 文獻；或重要臨床方法或成本的單一研究；或預後研究
Level 3	有對照組 (controlled study)		系統性回顧 Level 3 文獻；或不連續或缺乏公認標準驗證的研究	不連續或小族群的世代研究	其他臨床方法或成本的研究，包括敏感度(sensitivity) 分析
Level 4	病例系列	病例系列	對照病例研究 (case- control study)	病例系列	未分析敏感度
Level 5	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 團隊 Bob Philips and Chris Ball et al, since 1998)

Grades of Recommendation

A	consistent level 1 studies
B	consistent level 2 or 3 studies ***or*** extrapolations from level 1 studies
C	level 4 studies ***or*** extrapolations from level 2 or 3 studies
D	level 5 evidence ***or*** 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*



Modified from R Brain Haynes et al.: ACP Journal Club Nov/Dec 2006 | Vol 145 • Number 34;A8-A9.

THANK YOU FOR YOUR ATTENTION.