實證醫學研究之系統性回顧方法評論
—療效研究之品質評估

徐偉岸[1]　鍾國彪[2]

摘　要

評估研究的品質爲從事實證醫學研究，尤其是系統性回顧研究的關鍵步驟所在。回顧者若同時採用不同品質的原著研究，將會影響系統性回顧之最後結果，甚至導致完全相反的結論。當回顧眾多的原著文獻時，從選擇研究的類型開始，直到最後產生臨床醫療或進一步研究的推薦，每一階段都需要做研究之「品質評估」。品質評估的結果將影響到對該研究結果作推論的強度，以及最終在實證醫學上推薦的等級。研究之品質評估必須評估研究本身的內部效度及外部效度，方法上首先須訂出選取原著研究的最低品質要求標準。其次是應用品質檢查表及量表，來詳細審查所選出每一原著研究的細項品質，並釐清品質的差異對研究結果差異性的影響有多大。而療效研究的品質評估，主要聚焦在內部效度的部分。作者將提供有關療效研究之品質評估的方法及工具。至於非療效類的研究亦有其評估標準，本文只作概述介紹。

關鍵字：實證醫學、系統性回顧、品質評估、療效研究

壹、前　言

實證醫學 (Evidence-Based Medicine，EBM) 的興起相對於過往傳統西方醫學，到底是一種典範的轉移還是一時的流行？根據 David L. Sackett 的大作「Evidence-Based Medicine：How to Practice and Teach EBM」中指出：實證醫學的概念可回溯至十九世紀中大革命後的巴黎，當時臨床醫師如 Pierre Louis 拒絕相信權威，轉而尋求經由系統化的觀察病人結果作爲醫療依據。而到了近代，實證醫學的觀念被具體化，並於 1992 由加拿大 McMaster University 的 Gordon Guyatt 所領導的小組正式定名「實證醫學 (EBM)」，此後再藉由資訊科技的催化而成爲當代顯學。有趣的是，約莫同時期，自 1960 年代的醫學教育改革亦源於加拿大 McMaster University，強調以問題爲導向的學習 (Problem-Based Learning)。至此經由教育及再教育，臨床醫學的改革促使醫療工作者以「病人爲中心」的學習及照護，逐漸脫離對經驗醫學的信奉，強調以病人結果爲實證基礎的醫學，來取代權威性的專家之言。於是不論是公家（如英國 NHS R&D Center for Evidence-Based Medicine）或私人機構 (如 Cochrane Collaboration)、研究或推廣實證醫學的中心如雨後春筍般紛紛成立，並藉此建立實證醫學研究證據的強度等級 (level of evidence)，以供臨床醫療人員做爲是否要施行該

[1]輔仁大學醫學系講師；耕莘醫院內科醫師；臺灣大學醫療機構管理研究所博士生
[2]臺灣大學醫療機構管理研究所副教授
受文日期：2004 年 11 月 30 日　　修改日期：2005 年 3 月 22 日　　接受刊載：2005 年 5 月 4 日
通訊作者：鍾國彪　　台北市 112 徐州路 19 號 309 室

種診療之參考。

促使實證醫學由模糊的概念變成具體可行的科學，最重要的歷史進展首推 1940 年代 Archie Cochrane 建立隨機分派對照實驗 (randomized controlled trials) 的臨床實驗方法（梁， 2003）；其次是系統性回顧研究 (systematic reviews) 方法學的建立。 Deborah J. Cook 等人於 1998 年出版專書「Systematic Reviews：Synthesis of Best Evidence for Health Care Decisions」詳細介紹系統性回顧的步驟與方法，書中開宗明義指出系統性回顧是一種科學研究，目的是針對特定臨床問題，以事先計劃好的方法，採用明示可重製的標準 (explicit and reproducible criteria) 來搜集相關的原著研究，並加以嚴格評論及記錄，以釐清研究的偏差 (bias)，再將各個原著研究的結果綜合起來做定性 (qualitative systematic review) 或定量 (quantitative systematic review ，即 meta-analysis) 的總結。

作實證醫學系統性回顧研究時，所搜集研究的品質即代表該研究的證據強弱，若同時採用不同品質的研究，則會影響最後綜合分析的結果，甚至導致完全相反的結論，所以評估研究品質的方法確爲實證研究的重點所在 (Khan, 1996)。當回顧眾多的原著文獻時，從選擇研究的類型開始，直到最後產生臨床或研究的指南，每一階段都需要做「研究品質的評估」。本文將提供有關療效研究 (effectiveness studies) 之品質評估的背景知識，以及如何發展一套用來評估療效研究之品質的檢查表。而除療效的研究外，其他需要做品質評估的研究，還包括檢驗準確性 (test accuracy) 研究、質性研究以及健康經濟評估 (health economic evaluations) 等，本文將亦作概述。

貳、研究品質之評估

作系統性回顧研究時，必須作品質評估 (quality assessment) 之理由包括：探討造成研究結果差異性的原因，是否爲各研究的品質水準不同之故；以及便於最後能將研究結果依品質高低等級加以權重以作綜效分析 (meta-analysis)。品質評估的結果將影響該研究結果在實證醫學上作推論的強度 (level of evidence)，以及最後推薦的等級 (grade of recommendation)。譬如某設計不良之藥物隨機分派對照實驗（如未作雙盲、非眞正隨機分派），回顧者應將該實驗結果之證據強度由第一級降爲第二級，但若因回顧者未查覺此點而未予降級，或將之置於與其他設計良好之隨機分派對照實驗（第一級證據強度）同等權重一起分析後，所得結論可能由治療無效轉爲治療有效，因而給予臨床醫師相反的使用推薦。實際上，此藥物之眞正療效並未有定論，各實驗結果有的有效、有的無效，可能是因研究設計的品質差異所造成相反的結果，而非代表該藥眞正的療效。

關於「研究品質」究竟爲何物，眾家說法不同。通常在做品質評估時，需同時考慮文獻內容所呈現出該研究本身的「內部」及「外部」效度 (Verhagen, 1998；Deeks, 1998)。內部效度是指一個研究結果趨向「眞實」的程度，本身亦爲外部效度的先決條件。外部效度則是指一個研究結果能應用到此研究以外之一般情況的程度。因爲要解釋一個研究的結果爲何是如此，除了必須通盤考量其研究的設計、進行與分析方法（以上爲內部效度）之外，還須考量其研究的對象、是何種診療的介入，以及對結果的測量方法是否適當（以上爲外部效度），而這些研究的特點其實乃取決於最初「研究問題」是如何形成的。但一般說來，療效研究的品質評估，主要還是聚焦在「內部」效度的部分。

一、偏差與品質評估

偏差 (bias) 亦稱爲系統性錯誤 (systematic error)，指的是導致系統偏離「眞實結果」的傾向。在事後才來評估研究的偏差往往是困難且不可行的；所以，理論上，研究者應在事前就該儘量避免偏差的產生。但實際情形是：一般原著研究很難完全避開所有的偏差 (Sackett, 1979)。正因如此，在回顧這些研究文獻時，更需要有一套完整的品

表一　偏差的種類及避免偏差的研究設計

1. 選擇性偏差 因欲比較的各組對預後或對治療反應所產生的系統性差異。採隨機分派足夠數量的病人，並密封隨機分派碼，不讓病人知道自己屬於哪一組，可避免產生此類偏差。
2. 執行性偏差 計劃之外的照護介入所產生的系統性差異。採標準化照護，及對醫師及病人作雙盲設計（雙方皆不知病人身屬何種照護介入），可避免產生此類偏差。
3. 測量性偏差 因欲比較的各組對效果的測量標準不同所產生的系統性差異。對測量者及被測量者作雙盲設計（雙方皆不知此測量與何種照護介入有關），可避免產生此類偏差。
4. 耗損性（排除性）偏差 因參與者中途主動退出或被研究者排除（譬如對治療發生副作用）所產生的系統性差異。將這些參與者都納入研究並作敏感度分析，可避免產生此類偏差。

質評估步驟。

首先，要知道不同型態的偏差如何影響研究結果。回顧者必須仔細查看原著研究的內容，搜尋是否有證據顯示該研究者事前就運用一些方法來避免偏差的產生（表一）。譬如：唯有在不同比較組的基本特性都相近的前題下，推論某一治療的效果，才算具有良好效度（無選擇性偏差）的研究。因爲如果各組間基本特性差異太大，則無法確認最後效果上的差異是來自不同的治療介入還是來自病患原本不同的基本特性。由此原則可知，「觀察性研究」(如世代研究或案例對照研究)的效度會比「實驗性研究」低，主要是前者易發生這類「選擇性偏差」(selection bias) 的關係。兩者之間的差異在於實驗性研究可以藉由隨機分派研究對象至實驗組及對照組(randomization)，來平衡各組間已知或未知的干擾因子，而觀察性研究較不易做到這點。有一些實證研究指出對同一介入效果的估計，不論做觀察性研究或實驗性研究，其結果差異不大 (Benson, 2000)，但另一些實證研究則發現有明顯差異 (Sacks, 1982)。目前一般的共識是，無隨機分派的研究很難避開選擇性偏差所造成的偏離效果。

「執行性偏差」(performance bias) 指執行研究的過程中，出現計劃之外的介入因素，或研究中隱藏著未明示的其他介入因素所導致的偏差。這種偏差與研究本身未作「雙盲」(double blind) 設計有關（指研究者及研究對象皆對隨機分派的結果不知情）。此外，當研究結果是主觀性的測量時，也要設計讓研究對象及測量結果之評估者 (outcome assessor) 對診療介入是“雙盲”的， 才能避免「測量性偏差」(detection bias) 的產生。「耗損性偏差」(attrition bias) 指在研究追蹤過程中，因研究對象主動退出或被研究者排除，而導致追蹤時期的資料空白，進而影響效果的分析。所以好的研究報告，除了要明示研究對象的退出率或追蹤率外，最好採用所有曾「意圖治療」(intention to treat) 的對象來分析結果，以避免耗損性偏差。此外，爲補救追蹤時期的空白或遺漏資料，可採用敏感度分析 (sensitivity analysis) 來計算該情況可能影響效果的程度。

二、設定選取研究的最低品質要求標準

在作系統性回顧的方法中，應載明回顧者們決定要排除或選取該研究的條件，包括列出所能容許的最低品質標準，目的是讓回顧者能有效率地初步評估該研究設計是否適當。我們可以依照研

表二　療效研究之品質等級表

第一級　實驗性研究（隨機分派對照實驗，並密封分派碼）指研究者能有效控制某些研究狀況的研究，譬如能主動決定如何分派研究對象的方式（隨機或非隨機）。隨機分派對照實驗：追踪被隨機分派到實驗或對照組的所有研究對象，並比較兩組間效果的差異。隨機分派（並密封分派碼）可避開選擇性偏差對診療介入效果的影響。

第二級　擬實驗性研究
此研究由研究者控制如何分派研究對象至不同診療介入之組別，但未採用隨機分派並密封分派碼的方法。

第三級　對照觀察性研究
指該研究專門探討存在於不同診療介入間的本質差異，或探討研究對象間的不同暴露因子，以釐清診療介入或暴露因子對健康的影響效果爲何。可分爲下列幾種：
1.世代研究
追踪某群研究對象一段時間，比較那些已接受某種診療介入或暴露因子的對象與未接受者的後果有何差異（但此介入或暴露並非由研究者所指定）。
2.案例對照研究
比較已呈現某種後果（即“案例”組）及未呈現此種後果（即“對照”組）的研究對象，其所接受的診療介入或暴露因子有何差異。

第四級　無對照的觀察性研究，可分爲下列幾種：
1. 橫斷面研究
探討某一特定時點下某一特定族群，其疾病與其他研究變數間的關係。
2. 前－後比較研究
比較同一研究對象在接受某種診療介入前後有何異同。
3. 案例系列
僅描述不同案例接受同一診療介入情況及其後果，而未與控制組作任何比較。

第五級　專家意見

參考資料：Study quality assessment. 2001 CRD Report14

究問題的本質來製作一個依高低順序排列的研究品質等級表，再從中選定一個切點（即最低水準），在此切點水準以上的研究才納入系統性回顧。當研究問題是：比較相同情況下，不同介入治療的效果時（即療效之研究），最佳品質的研究設計應是「雙盲，隨機分派對照實驗」(double-blind, randomized controlled trial)，因爲此設計可防止發生選擇性偏差。次佳的研究設計則是「擬實驗性研究」(quasi-experiment)，亦即非隨機分派之對照實驗；再其次是觀察性研究 (observational study)。在作系統性文獻回顧時，若無法選取足夠份量的最佳研究（如雙盲，隨機分派對照實驗），則考慮加入次佳的研究（如擬實驗性研究或觀察性研究），直到回顧者所能接受之最低水準以下之研究則予排除。表二列出常用的療效研究之品質等級表以供參考。

三、運用品質評估的工具深入分析每一篇已選取文獻的研究品質

一旦依最低的品質標準選取出所需回顧的原著文獻之後，接著將深入分析每一篇研究的執行品質；因爲就算是依同一標準所選取的研究（如都是隨機分派到對照或實驗組），各研究間的執行品質仍有差異。譬如，隨機分派不完整，介入因

表三　發展品質評估工具的步驟

1. 定義品質的內容結構
   考慮幾個品質的構面：如，內部效度、外部效度及統計分析。
2. 定義品質評估的範圍及目的
   考慮所欲回顧的研究設計的型態及研究問題爲何。
3. 發展品質評估的工具（檢查表）
   將相關的品質單項予以群組化，並發展一套配分的系統。品質單項可依下列類型予以群組化：
   (1) 通用的：適用於一般研究中都會提及有關研究設計重點的單項。
   (2) 特定的：適用於某回顧領域內有關研究品質的單項。又可分爲方法學上的及臨床的兩組。
   (3) 質性描述的（此項可以不予配分）。
4. 評估此工具的各方面測量能力
   此工具在正式應用於大量文獻回顧前，需先作小型先驅研究，來確認其信度。

子的比較不適當，效果測量未作“雙盲”，追蹤時間太短，效果未清楚定義或其測量方式不可靠，乃至錯誤的統計方法等等不良的執行品質，都會影響系統性回顧的最後推論 (Khan, 1996；Kunz, 1998；Schulz, 1995)。

作系統性回顧時，最好將不同標準下所選取出的研究資料，分開來作資料整合。另一種分析的方法，是將「研究品質」設定爲一自變數，放入綜效迴歸分析 (meta-regression) 中，來檢驗不同的品質水準對研究效果的影響。注意，當各研究效果的差異性甚大時，回顧者應予品質較佳的研究較高的權重。依品質水準的高低賦予各研究不同的權重後，就可以進行綜效分析了 (Detsky, 1992)，這些證據的品質標準，也將決定該系統性回顧的推薦強度。

常見的品質評估工具的內容包括個別品質單項 (individual quality items)、品質檢查表 (quality checklist)及品質量表 (quality scale)。個別品質單項應涵蓋研究設計，執行及分析三個面項；這些品質項目可以列成一套檢查表，以便回顧者能系統性地檢查每一選取出來的研究文獻 (Khan, 2001)。若賦予檢查表每一項目一個分數，就成爲品質的分數量表。目前已有多種套裝的品質評估量表發展出來 (Miher, 1995)，但大多只適用於單一研究形式（實驗性或觀察性）。如果要同時回顧實驗性及觀察性研究，可以選擇分開的兩種品質評估量表，或自行合併爲一份品質評估量表。

設計一份品質評估表可依據表三所列的原則訂出測量的尺度。對同一研究，採用不同的評估量表可能會得出不同的品質分數，進而影響綜效分析的結果，尤其是包括外部效度的品質單項時，影響極大。所以，在作研究品質之評估時，最好採用個別品質單項分數而非總分來分析。

品質評估的方法應明載於系統性回顧方法中，且檢查表或分數量表應列入回顧資料的一部分。理想上，文獻回顧者可依表三的方式自訂品質評估量表，但只要事前決定所要回顧的研究類型及品質內容，回顧者通常會採用套裝的品質評估量表，因這些工具已足夠涵蓋各種方法學上所強調的研究設計，回顧者再依其特別需求來作增修即可。

文獻回顧者在評估每一品質單項時，極容易陷於主觀。因此，在品質檢查表中，回顧者應說明是如何評估該品質單項。同時，此檢查表應先作小規模先驅研究以評估其信度。如先驅研究顯示信度差，補救之道，則由回顧者採用統一明確的編碼系統來勾選各品質單項，可提高回顧者本身的一致性 (intra-rater agreement)。至於是否一定

要彌封原著之作者姓名、研究機構及期刊名稱，不讓回顧者知道，以免影響評分高低，尚未有定論 (Jadad, 1996；Clark, 1999)。一般認爲，在明示狀態下，分別由兩位以上的回顧者獨自作品質評估，即具有一定的公信力。

四、 療效研究的品質評估

我們可依表二之品質標準排列出療效研究的相對等級關係。需注意，隨機分派對照實驗唯有在"完整執行"時才列爲最高品質等級（第一級）。完整的執行包括：隨機分派足夠數量的參與者、密封隨機分派碼、不讓參與者與研究者知道隨機分派碼（雙盲）、接近百分之百的追蹤率、及採用「意圖治療」來分析結果 (Khan, 1996；Schulz, 1995)。如果針對某主題，經文獻搜尋後，發現缺乏完整執行的隨機分派對照實驗，則應優先考慮完整執行的擬實驗性研究或觀察性研究（品質高於執行不完整的隨機分派對照實驗）。因此，研究品質的等級順序並非一成不變的。

關於療效研究中的實驗性研究或觀察性研究現已有眾多品質評估標準可供參考，茲簡介如下：

(一) 實驗性研究

實驗性研究以完整的執行的隨機分派對照實驗爲最高標準。回顧隨機分派對照實驗時，低品質研究可能會產生與高品質研究相反的結果。有相當多的文獻提出如何評估此類研究的品質標準；著名的例子有 Verhagen 等人採用 Delphi 方法自 206 項品質標準中所挑出 9 項有共識的標準。這個標準表也可適用於評估擬實驗性研究 (Verhagen, 1998)。另 Jadad 等人亦運用類似方法發展出 5 項評估隨機分派對照實驗之標準，包含是否明示爲隨機分派研究、隨機分派的方法是否正確、是否明示爲雙盲研究、雙盲設計是否正確、及是否明示病人退出或脫離研究之情況，即所謂 Jadad score，簡易實用，且其效度已經研究確認 (Jadad, 1996)。

(二) 觀察性研究

觀察性研究也可自成一階層。世代研究（cohort study，比較同一時點，接受不同介入的各組對象）的效度通常比歷史性對照研究 (historical control study) 要高，這意謂不同時點的兩組對象比同一時點的兩組對象存在有更大的差異。前瞻性的世代研究又比回溯性的世代研究更不易產生誤謬，這是因爲前瞻性有計劃的收集資料比回溯性收集更趨完整可信，且前瞻性選擇病人較不易影響最後效果。

案例對照研究 (case-control study) 較世代研究更易發生誤謬。「配對」(matching) 是一個用來平衡兩組間或個案間的干擾因子 (confounding factors) 的方法。但注意，過度配對可能導致治療效果被低估。其他避開誤謬的方法包括：治療介入的種類與效果測量互不公開，案例或對照組的治療介入方式也不公開。

第三種研究設計是前－後設計 (before-after design)，指只比較同一組對象介入前後的結果。另一種是時間系列設計(time series design)，指介入前後一段時間內重覆測量幾次。時間系列設計比簡單的前－後設計提供可信度更高的資訊。 有時候，我們搜尋不到有關不同介入的比較性研究，而只能搜尋到各個不同介入的案例報告 (case report)，這些案例報告應只適合拿來產生未來研究的假說而不應拿來互相比較。

用來評估流行病學研究（尤其是有關預後及危險因子之研究）的品質檢查表，因方法學上近似，通常可以改良成爲評估觀察性研究的檢查表及量表 (Levine, 1994；Fleiss, 1991)。

五、其他非療效研究的品質評估

(一) 檢驗之準確性研究

針對疾病篩檢及診斷的研究，一般都著重於評估檢驗的準確性，也就是某項檢驗能正確診斷或正確排除某病的能力。關於這類研究的品質評估與前述療效研究甚爲不同，較佳的研究設計應是前瞻性研究，對所有合適的研究對象都採用某一檢驗方式，再用「參考標準」(reference standard) 的檢驗來確定或排除研究對象患有某病，兩者相比較，才能評估該檢驗的準確性 (Mulrow, 1989)。

檢驗準確性研究的品質評估分級方法之一，是依據此類研究易犯的誤謬，如受試者樣本不足、錯用「參考標準」的檢驗、未作盲目設計及僅做案例控制研究等，這些都會影響檢驗準確性的評估 (McAlister, 1999)。

值得注意的是：雖然我們仍可以用上述療效研究的分級方法（如表二）來評估某檢驗用來篩檢疾病的存活效果，但其中有所謂「時間偏差」(time bias) 的存在。特別是在世代研究中，當下被檢驗出有病的族群，並不會因而增加此後之存活時間，只有那些在疾病的早期被診斷出來而又經過適當的治療者，才會可能有較長的存活時間；亦即「檢驗」本身並無增加存活時間的效果。

(二) 質性研究

愈來愈多關於健康及社會照顧的實證是來自質性研究，以致作系統性文獻回顧時絕不能忽視此點，需質性與量性研究並重。雖然質性研究的誤謬較不易控制，但仍有結構性的方法可作品質評估。不過，雖然研究者有心建立一套質性研究的品質標準，至今仍無統一分級的方法，主要是對「品質」如何測量有異議 (Buchanan, 1992)。

撇開各種爭論，質性研究的效度與信度評估方法已在正式發展中 (Popay, 1998；BSA Medical Society Group, 1996)。Popay 及其同事指出，在進入詳細評估方法學上是否健全之前，應先評估是否用對了方法，亦即先確認此問題是否適用質性研究 (Popay, 1998)。但在其他專家所發展的評估架構中卻常不提此點，因爲怕會產生爭議。事實上，眞正爭議點在於“如何去進行評估”（如何決定此問題是否適用質性研究），而光這一點，就常未能在評估架構中說明清楚。諸如此類的不確定性，導致現階段仍無法針對質性研究產生一致的品質分級方法。

(三) 經濟評估研究

實證研究的文獻回顧愈來愈重視效率 (efficiency)，亦即所謂經濟評估。能被選入系統性回顧的經濟評估研究，須對多種（至少兩種）介入的成本及臨床效果同時作清楚的分析。一旦被選入回顧，每一研究品質必需接受有系統的嚴格評論 (Drummond, 1997)。

在作經濟評估的品質評估時要注意區分該評估的類型是實驗性質 (trial-based) 或模型性質 (model-based)。如果是實驗性質的經濟評估，將只用到單一的實證來源，所以品質評估的重點是研究設計的「內部效度」。如果是模型性質的經濟評估，則會用到多方的實證來源，其品質評估的重點包括：文獻回顧的品質、模型參數的求得方式、模型是否具代表性、及敏感度分析等。一般模型性質的經濟評估會利用到「決策樹」(decision trees) 或「馬可夫模型」(Markov models) 來作臨床決策：譬如，病人在追蹤時期的某一時點可能由某健康狀態（無病或有病）轉變爲另一狀態（痊癒、殘障或死亡）。馬可夫模型更能將疾病的自然史及健康狀態的轉換機率一併考慮 (Briggs, 1998)。

參、結　語

實證醫學雖已成爲當代臨床醫學之主流，其擁護者常借用孔恩 (Thomas Kuhn) 的科學遽變理論，宣稱實證醫學是一種典範的轉移 (paradigm shift)。但它本身卻面臨許多反對者嚴苛的質疑，包括缺乏實證證明「實證醫學」眞能改善病人照護結果。最近 Sehon and Stanley 卻揚棄孔恩，改以昆恩 (W.V. Quine) 的信念網絡理論 (web of belief)，肯定實證醫學是在此網絡中延伸經驗醫學的觀察，以更爲嚴謹，更爲科學的方法去瞭解介入療效的問題 (Sehon & Stanley, 2003)。所謂更嚴謹、更科學的方法其實指的就是能避開偏差，讓結果更趨於眞實。

本文限於篇幅，無法介紹實證醫學之全貌，主要是針對其中的方法學部分作一回顧。誠如本文前言強調：系統性回顧是一種嚴謹的科學研究，而如何評估原著研究的品質，以釐清可能產生的偏差，實爲影響系統性回顧研究結果之關鍵所在，故其相關知識值得介紹給醫護工作者。

經本文之回顧，可以得知：研究之品質評估必

須評估研究本身的「內部效度」及「外部效度」。第一步是先訂出選取原著研究的最低品質要求。第二步是應用品質檢查表及量表，詳細審查所選出每一原著研究的品質細項，並釐清研究的品質差異對研究結果差異性的影響有多大。品質檢查表及量表可以自行研發或採用套裝量表。檢查表及量表之品質單項應包括一般通用及特殊專門領域相關的議題。而採用品質單項的分數比採用總分更能分析出品質對該研究的影響力。

肆、參考文獻

梁繼權 (2003)．臨床醫學模式的典範轉移與實證醫學的發展．*臺灣醫學*，*7*(4)，531-534。

Benson, K., & Hartz, A. (2000). A comparison of observational studies and randomized controlled trials. *New England Journal of Medicine*, *342*, 1878-1886.

Briggs, A., & Sculpher, M. (1998). An introduction to Markov modeling for economic evaluation. *Pharmacoeconomics*, *13*, 397-409.

BSA Medical Society Group. (1996). Criteria for the evaluation of qualitative research papers. *Medical Sociology News,* 22.

Duchanan, D. R. (1992). An uneasy alliance: combining qualitative and quantitative research methods. *Health Education Quarterly*, *19*, 117-135.

Clark, H. D., Wells, G. A., & Huet, C. (1999). Assessing the quality of randomized trials: the reliability of the Jadad scale. *Controlled Clinical Trials*, 20, 448-452.

Mulrow C. D., & Cook D. J. (1998). *Systematic Reviews: Synthesis of Best Evidence for Health Care Decisions*. Philadelphia, Pennsylvania: American College of Physicians.

Deeks, J., & Altman, D. (1998). Inadequate reporting of controlled trials as short reports. *Lancet*, 352, 1908.

Detsky, A. S., Naylor, C. D., O'Rourke, K., McGeer, A. J., & L'Abbe, K. A. (1992). Incorporating variations in the quality of individual randomized trials to meta analysis. *Journal of Clinical Epidemiology*, *45*, 255-265.

Drummond, M. F., O'Brien, B., Stoddard, G., & Torrance, G. (1997). *Methods for the economic evaluation of health care programs. 2nd ed.* Oxford University Press.

Fleiss, J. L., & Gross, A. J. (1991). Meta-analysis in epidemiology, with special reference to studies of association between exposure to environmental tobacco smoke and lung cancer: a critique. *Journal of Clinical Epidemiology*, *44*, 127-139.

Jadad, A. R., Moor, R. A., & Carroll, D., et al. (1996). Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Controlled Clinical Trials*, *17*, 1-12.

Khan, K. S., Daya, S., & Jadad, A. R. (1996). The importance of primary studies in producing unbiased systematic reviews. *Archives of Internal Medicine*, *156*, 661-666.

Khan, K. S., Riet, G., Popay, J., Nixon, J., & Kleijen, J. (2001). Study quality assessment. In Undertaking Systematic Reviews of Research on Effectiveness. *CRD Report*, *4*(5), 6-8.

Kunz, R., & Oxman, A.D.(1998). The unpredictability paradox: review of empirical comparisons of randomized and nonrandomized clinical trials. *British Medical Journal*, *317*, 1185-1190.

Levine, M., Walters, S., & Lee, H. et al. (1994). User's guide to the medical literature. IV: How to use an article about harm. *Journal of American Medical Association*, *271*, 1615-1619.

McAlister, F. A., Straus, S. E., & Sackett, D. L. (1999). Why we need large, simple studies of

the clinical examination: the problem and a proposed solution. *Lancet*, *354*, 1721-1724.

Miher, D., Jadad, A. R., Nichol, G., Penman, M., Tugwell, P., & Walsh, S. (1995). Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklist. *Controlled Clinical Trials*, *16*, 62-73.

Mulrow, C. D., Linn, W. D., & Gaul, M. K. (1989). Assessing quality of a diagnostic test evaluation. *Journal of General Internal Medicine*, *4*, 288-295.

Popay, J., Roger, A., & Williams, G. (1998). Rationale and standards in the systematic review of qualitative literature in health service research. *Qualitative Health Research*, *8*, 341-352.

Sackett, D. L. (1979). Bias in analytical research. *Journal of Chronic Disease*, *32*, 51-63.

Sackett, D. L., Straus, S. E., Richardson, W.S., Rosenberg, W., & Haynes, R. B. (2000). *Evidence-Based Medicine: How to Practice and Teach EBM*. New York, NY: Churchill Livingstone.

Sacks, H. S., Chalmers, T. C., & Smith, H. J. (1982). Randomized versus historic assignment in controlled clinical trials. *American Journal of Medicine*, *72*, 233-240.

Schulz, K., Chalmers, I., Hayes, R. J., & Altman, D. G. (1995). Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled clinical trials. *Journal of American Medical Association*, *273*, 408-412.

Sehon S. R., & Stanley D. E. (2003). A philosophical analysis of the evidence-based medicine debate. *BMC Health Services Research*, *3*(14), 1-10.

Verhagen, A. P., de Vel, H. C., & de Bie, R. A. et al. (1998). The delphi list, a criteria list for quality assessment of randomized clinic trials for conducting systematic reviews developed by delphi consensus. *Journal of Clinical Epidemiology*, *51*, 1235-1241.

The Journal of Health Science: July 2005, Vol. 7, No. 3. p307~p316

Assessing the Quality of Effectiveness Studies for Systematic Review in Evidence-Based Medicine

Wei-An Hsu[1] Kuo-Piao Chung[2]

Abstract

Quality assessment of selected studies is critical to a systematic review, a scientific method applied to evidence-based medicine. Studies with varying quality will significantly impact a final conclusion and, eventually, can lead to incorrect recommendations. In systematic review, quality assessment is essential at each stage from initial selection of literature to final clinical practice and research recommendations. Assessing study quality is to determine internal and external validity. First, reviewers should establish minimal quality requirements for inclusion and exclusion of studies exploring related topics. Then reviewers should develop and apply a quality check list and/or quality scale to examine in detail the quality of each study to clarify the impact of study quality on study outcomes. In effectiveness studies, the principal criteria for study quality are primarily concerned with internal validity. This work introduces a methodology and tools to assess quality of effectiveness studies. This work also provides a brief overview for studies other than effectiveness.

Key words: Evidence-based medicine, systematic reviews, quality assessment, effectiveness studies

[1] Instructor, College of Medicine, Fu-Jen Catholic University; MD, Department of Internal Medicine, Cardinal Tien Hospital; Doctoral student, Institute of Health Care Organization Administration, National Taiwan University.
[2] Associate Professor, Institute of Health Care Organization Administration, National Taiwan University.
Received: Nov. 30, 2004 Revised: Mar. 22, 2005 Accepted: May 4, 2005
Address Correspondence to: Kuo-Piao Chung Rm 309, No. 19, Suchow Rd., Taipei 112, Taiwan (R.O.C.)