

근거수준별 문헌의 질 평가 도구

임 선 미¹ · 신 인 순^{2*} · 이 선 희² · 서 경 화¹ · 정 유 민² · 장 지 은² | ¹대한의사협회 의료정책연구소, ²이화여자대학교 의과대학 예방의학교실

Tools for assessing quality and risk of bias by levels of evidence

Sun Mi Lim, PhD¹ · Ein Soon Shin, PhD^{2*} · Sun Hee Lee, MD² · Kyung Hwa Seo, MBA¹ · Yu Min Jung, RN² · Ji Eun Jang, BS²

¹Research Institute for Healthcare Policy, Korean Medical Association, ²Department of Preventive Medicine, Ewha Womans University School of Medicine, Seoul, Korea

*Corresponding author: Ein Soon Shin, E-mail: shin1121@ewha.ac.kr

Received January 30, 2011 · Accepted February 17, 2011

Tools for assessing methodological quality or risk of bias in randomized controlled trials (RCTs) and non-randomized studies (NRS) were reviewed. The van Tulder scale and Cochrane's assessment of risk of bias are the two most useful methodological quality evaluation tools for RCTs. Cochrane's tool includes sequence generation, allocation of sequence concealment, blinding, incomplete outcome data, selective outcome reporting, and other potential sources of bias. The Cochrane Collaboration Group recommends the Downs and Black instrument and the Newcastle-Ottawa Scale for evaluating the quality of NRS. In conclusion, this study offers useful information to physicians about tools for assessing the quality of evidence in clinical guidelines. Further research is needed to provide an essential core for evidence-based decision making regarding levels and/or grades of recommendations.

Keywords: Jadad's scale; van Tulder scale; Cochrane's tool; Downs and Black instrument; Newcastle-Ottawa Scale

서 론

사람의 생명을 다루는 보건의료 분야에서 사용되고 있는 다양한 근거 자료들을 체계적으로 요약하고 그 근거의 질을 종합적이고 구체적인 기준에 의해 평가[1-3] 하는 것은 근거기반 의사결정[4-9] 및 근거중심의학[10-14]의 핵심적 방법론에 속한다. 특히 근거의 질을 평가하는 과정은 임상현장에서 과학적인 진료문화[15]를 정착시키는데 핵심적인 역

할[16-19]을 함으로써 궁극적으로 의료의 질[20]을 높일 뿐만 아니라, 근거평가 결과에서 불필요하거나 잘못된 근거 자료들을 판별해내고 이를 임상현장에 적용함으로써 의료비를 절감하는 잠정적인 효과까지 기대 할 수 있다. 아울러 최근 국내에서도 높은 관심의 대상이 되고 있는 임상진료지침 개발의 핵심 구성요소이며 임상진료지침의 질에 결정적인 영향을 미치기도 한다는 점에서 중요한 방법론적 이슈가 되고 있다.

특히 국내 문헌 혹은 근거의 대부분이 비무작위 연구

(non-randomized study, NRS)임에도 불구하고 이에 대한 질 평가 도구가 부재한 실정임을 감안할 때, 무작위배정 비교임상시험(prospective randomized controlled double blind clinical trial, RCT) 뿐만 아니라 비무작위 연구 근거들에 대하여 우리나라 현실에 맞는 한국형 진료지침 제작에 필요한 근거수준별 질 평가 도구[21-22]를 발굴하여 적용 타당성에 대한 체계적인 검토를 하는 일은 꼭 필요하다. 이러한 필요성과 의미를 고려할 때, 근거수준별로 문헌의 질 평가에 활용할 수 있는 국제적 도구들을 검토하고 이중 국내에서 주로 생산되고 있는 근거의 실정에 부합할 수 있는 도구 및 활용 방법을 제시하는 것은 근거기반 의료문화를 정착시켜나가는데 있어 중요한 연구 작업이라고 할 수 있겠다.

RCT 연구에 사용되는 근거 및 문헌의 질 평가 도구

RCT 연구에 사용되는 근거 및 문헌의 질 평가 도구들 중에서 Jadad, van Tulder, Cochrane, Newell's, Scottish Intercollegiate Guidelines Network (SIGN), 그리고 National Institute for Health and Clinical Excellence (NICE)에서 사용하고 있는 평가 도구들은 타당도가 검증된 척도 혹은 체크리스트로서 국제적으로 많이 쓰이고 있다. RCT의 내적타당도를 향상시키기 위해 실험적 근거에 기초하여 개발된 consolidated standards of reporting trials (CONSORT) 체크리스트 25개 항목들을 기준으로 하여 이상의 6개 RCT 평가 도구에서 활용하고 있는 구성요소들을 비교해 본 결과 SIGN 체크리스트가 RCT 문헌평가에 있어서 가장 많은 부분을 포괄적으로 평가하고 있음을 알 수 있다(Table 1).

국·내외에서 많이 사용하고 있는 대표적인 평가 도구는 Jadad 척도와 코크란 연합의 평가방법이다. Jadad 척도는 평가 문항이 간단하고, 평가하기 쉬운 장점이 있어 그동안 국내·외에서 많이 사용되어 왔지만, RCT 연구 평가의 가장 중요한 항목인 무작위배정 순서의 할당 은닉에 대한 평가 항목을 포함하고 있지 않기 때문에 코크란 연합에서는 권장하

지 않고 있는 평가 도구이다. 코크란 연합의 체계적 문헌고찰 방법론에서 권장하고 있으면서 RCT 방법론에 대한 간편하고 타당한 평가 도구로는 van Tulder 척도가 있다.

1. Jadad 척도

Jadad 척도는 Oxford quality scoring system이라고도 알려져 있으며, RCT 문헌을 평가하기 위한 척도이며 총 5문항으로 구성되어 있다[23]. 즉, 무작위배정에 관한 평가문항이 2개, 눈가림에 관한 평가문항이 2개, 탈락에 관한 평가문항이 1개이다. 평가방법은 RCT 문헌에서 무작위화를 언급하면 1점, 적절한 무작위법을 포함하고 있으면 1점이 추가되며, 잘못된 무작위 평가는 -1점의 점수를 주게 되고, 총체적으로 0-2점의 점수분포를 갖는다. 눈가림에 대한 평가는 이중눈가림이 언급된 경우 1점, 적절한 이중 눈가림법을 포함하고 있으면 1점이 추가되며, 잘못된 이중눈가림을 표현하고 있으면 -1점의 점수를 주게 되고, 무작위배정에 관한 평가와 마찬가지로 최대 2점의 점수를 부여할 수 있다. 마지막으로 탈락에 관한 내용이 문헌에 언급되었을 경우 1점의 점수를 부여할 수 있다. 이렇게 5가지 항목에 대해 각 1점씩 점수를 줄 수 있으며, 총 5점 만점으로 평가한다. 최종 문헌의 질 평가는 평가점수가 0-2점 사이는 문헌의 질이 낮은 것으로, 3-5점 사이는 문헌의 질이 높은 것으로 평가한다(Table 2).

2. van Tulder 척도

van Tulder 척도[24]는 코크란 연합에서 권장하고 있는 RCT 문헌의 질 평가에 적합한 도구 중의 하나이다. 본 평가 도구는 무작위 적절성, 치료배정은닉, 기본특성의 유사성, 환자 눈가림, 치료제공자 눈가림, 관찰자 눈가림, 동시 중재, 규정 준수, 탈락/탈락률, 결과평가 시점, 그리고 무작위할당 대로의 분석과 같은 11개 구성요소로 평가하도록 되어 있다.

평가방법은 각각의 항목에 대하여 ‘예’, ‘아니오’, ‘알 수 없음’ 중 하나를 선택하도록 되어 있는데, 5개 이상의 평가항목에 대한 기준을 만족하는 경우(5점 이상) 문헌의 질이 높은 것으로 평가하고 있다(Table 3).

Table 1. Quality assessment tools of randomized controlled trials

CONSORT checklist	Jadad [23]	van Tulder [24]	Cochrane [1]	Newell's [25]	SIGN [26]	NICE [27]
1. Title and abstract						
• Identification as a randomized trial in the title						
• Structured summary of trial design, methods, results, and conclusions						
2. Introduction						
• background and objectives		√		√	√	
3. Methods						
• Trial design					√	√
• Participants					√	√
• Interventions		√		√	√	√
• Outcomes		√				√
• Sample size					√	
• Randomization	√	√		√	√	√
• Sequence generation			√			
• Allocation concealment		√	√	√	√	√
• Implementation						
• Blinding	√	√	√	√	√	√
• Statistical methods						
4. Results						
• Participant flow /numbers analysed/ drop outs	√	√	√	√	√	√
• Recruitment and follow-up			√	√	√	√
• Baseline data		√	√		√	
• Outcomes and estimation			√	√	√	√
• Ancillary analyses					√	√
• Harms			√		√	√
5. Discussion						
• Limitations			√			
• Generalisability						
• Interpretation						
6. Other information						
• Registration						
• Protocol						
• Funding					√	

CONSORT, consolidated standards of reporting trials; SIGN, Scottish Intercollegiate Guidelines Network; NICE, National Institute for Health and Clinical Excellence.

3. 코크란 연합의 질 평가 방법(Cochrane's assessment of risk of bias)

코크란 연합은 RCT 근거문헌의 평가를 위하여 6개 영역 즉, 순서 생성, 배정 은닉, 눈가림, 불완전한 결과변수, 선택적 결과보고, 그리고 타당도를 위협하는 기타 잠재적 편견으로 구분하여 평가하고 있다[1]. 평가방법은 각 영역별로 '예', '아니오', '분명하지 않음'으로 평가하도록 되어 있으며, 평가자는 6개의 개별 영역별로 yes, no, unclear에 대한 세부 기준에 따라 판단하고 평가하도록 되어 있다. '예'는 낮은 위

험편견을, '아니오'는 높은 위험편견을, 그리고 '분명하지 않음'은 정보가 불충분한 경우를 의미한다(Table 4).

NRS 연구에 활용되는 근거 및 문헌의 질 평가도구

Deeks 등[21]은 비무작위 연구의 방법론적 질 평가도구에 대한 체계적 문헌고찰 결과 213개의 비무작위 연구 평가 도구를 발굴한 바 있으며, 이때 역학연구, 중재 부작용에 대

Table 2. Jadad scale

Item	Maximum points	Description
Randomization	2	<ul style="list-style-type: none"> • 1 point if randomization is mentioned • 1 additional point if the method of randomization is appropriate • Deduct 1 point if the method of randomization is inappropriate (minimum 0)
Blinding	2	<ul style="list-style-type: none"> • 1 point if blinding is mentioned • 1 additional point if the method of blinding is appropriate • Deduct 1 point if the method of blinding is inappropriate (minimum 0)
An account of all patients	1	<ul style="list-style-type: none"> • The fate of all patients in the trial is known. If there are no data the reason is stated

From Jadad AR, et al. Control Clin Trials 1996;17:1-12, Appendix with permission from Elsevier [23].

Table 3. van Tulder scale [24]

Criteria		
A	Was the method of randomization adequate?	Yes/no/don't know
B	Was the treatment allocation concealed?	Yes/no/don't know
C	Were the groups similar at baseline regarding the most important prognostic indicators?	Yes/no/don't know
D	Was the patients blinded to the intervention?	Yes/no/don't know
E	Was the care provider blinded to the intervention?	Yes/no/don't know
F	Was the timing of the outcome assessment in all groups similar?	Yes/no/don't know
G	Were co-interventions avoided or similar?	Yes/no/don't know
H	Was the compliance acceptable in all groups?	Yes/no/don't know
I	Was the drop-out rate described and acceptable?	Yes/no/don't know
J	Was the timing of the outcome assessment in all groups similar?	Yes/no/don't know
K	Did the analysis include an intention-to-treat analysis?	Yes/no/don't know

한 조사 연구, 환자-대조군 연구, 그리고 통제되지 않은 연구들은 조사 대상에서 제외되었다. Deeks 등에 의하여 개발된 NRS 평가 도구들을 비교·평가하기 위한 12개 핵심영역의 구체적인 기준 및 세부항목들은 Table 5와 같다.

본 연구에서는 Downs and Black 도구, Newcastle-Ottawa 척도, Reish, Thomas, 그리고 Zaza 평가 도구들의 구성 항목들을 Deeks 등에 의해 개발된 NRS 평가 기준 및 세부항목들과 비교해 본 결과 Downs and Black 도구 및 Reisch 도구가 NRS 평가 기준에 가장 부합되는 유용한 도구임을 알 수 있었다(Table 5). 코크란 연합에서는 비무작위 연구의 질 평가를 위한 유용한 도구로 Downs and Black 도구를 그리고 환자-대조군 연구의 질 평가 도구로는 Newcastle-Ottawa 척도를 권장하고 있다.

1) Downs and Black 도구

Downs and Black 도구는 역학 연구의 기본원칙, 연구

설계의 검토, 그리고 RCT 평가에 대한 기존 체크리스트 등을 기반으로 개발된 것이다[28]. 연구의 전반적 기술 보고에 관한 항목 10개, 외적타당도에 관한 항목 3개, 내적타당도의 비뚤림에 관한 항목 7개, 내적타당도의 혼란요인 및 선택 비뚤림에 관한 항목 6개, 그리고 임상적으로 중요한 효과크기에 관한 항목 1개 즉, 5개 척도 27개 세부항목으로 구성되어 있는 평가 도구이다(Table 6).

평가방법은 각 항목에 0점 또는 1점을 줄 수 있으나, 예외적으로 연구의 전반적 기술보고에 관한 1개 항목은 최대 2점까지 줄 수 있고, 효과크기의 항목은 최대 5점까지 줄 수 있어, 최대점수는 32점이다. Downs and Black 도구는 타당도와 신뢰도 측면에서 질이 높은 평가도구이다. 그러나 환자-대조군 연구의 질을 평가하는 데에는 적합하지 않은 척도이다.

2) Newcastle-Ottawa 척도

Newcastle-Ottawa 척도는 비무작위 연구인 대부분의

Table 4. Cochrane's assessment of risk of bias [1]

Item		Criteria
Sequence generation	Yes	referring to a random number table, using a computer random number generator, coin tossing, shuffling cards or envelopes, throwing dice, drawing of lots, minimization*
	No	sequence generated by odd or even date of birth, sequence generated by some rule based on date (or day) of admission, sequence generated by some rule based on hospital or clinic record number
Allocation concealment	Yes	central allocation (including telephone, web-based and pharmacy-controlled randomization), sequentially numbered drug containers of identical appearance, sequentially numbered, opaque, sealed envelopes
	No	using an open random allocation schedule (e.g. a list of random numbers), assignment envelopes were used without appropriate safeguards, alternation or rotation, date of birth, case record number, any other explicitly unconcealed procedure
Blinding	Yes	<ul style="list-style-type: none"> No blinding, but the review authors judge that the outcome and the outcome measurement are not likely to be influenced by lack of blinding; Blinding of participants and key study personnel ensured, and unlikely that the blinding could have been broken; Either participants or some key study personnel were not blinded, but outcome assessment was blinded and the non-blinding of others unlikely to introduce bias
	No	<ul style="list-style-type: none"> No blinding or incomplete blinding, and the outcome or outcome measurement is likely to be influenced by lack of blinding; Blinding of key study participants and personnel attempted, but likely that the blinding could have been broken; Either participants or some key study personnel were not blinded, and the non-blinding of others likely to introduce bias
Incomplete outcome data	Yes	<ul style="list-style-type: none"> No missing outcome data; Reasons for missing outcome data unlikely to be related to true outcome Missing outcome data balanced in numbers across intervention groups, with similar reasons for missing data across groups; For dichotomous outcome data, the proportion of missing outcomes compared with observed event risk not enough to have a clinically relevant impact on the intervention effect estimate; For continuous outcome data, plausible effect size among missing outcomes not enough to have a clinically relevant impact on observed effect size; Missing data have been imputed using appropriate methods
	No	<ul style="list-style-type: none"> Reason for missing outcome data likely to be related to true outcome, with either imbalance in numbers or reasons for missing data across intervention groups; For dichotomous outcome data, the proportion of missing outcomes compared with observed event risk enough to induce clinically relevant bias in intervention effect estimate; For continuous outcome data, plausible effect size among missing outcomes enough to induce clinically relevant bias in observed effect size; 'As-treated' analysis done with substantial departure of the intervention received from that assigned at randomization; Potentially inappropriate application of simple imputation
Selective outcome reporting	Yes	<ul style="list-style-type: none"> The study protocol is available and all of the study's pre-specified (primary and secondary) outcomes that are of interest in the review have been reported in the pre-specified way; The study protocol is not available but it is clear that the published reports include all expected outcomes, including those that were pre-specified
	No	<ul style="list-style-type: none"> Not all of the study's pre-specified primary outcomes have been reported; One or more primary outcomes is reported using measurements, analysis methods or subsets of the data (e.g., subscales) that were not pre-specified; One or more reported primary outcomes were not pre-specified (unless clear justification for their reporting is provided, such as an unexpected adverse effect); One or more outcomes of interest in the review are reported incompletely so that they cannot be entered in a meta-analysis; The study report fails to include results for a key outcome that would be expected to have been reported for such a study
Other potential threats to validity	Yes	<ul style="list-style-type: none"> Had a potential source of bias related to the specific study design used; or Stopped early due to some data-dependent process (including a formal-stopping rule); or Had extreme baseline imbalance; or Has been claimed to have been fraudulent; or Had some other problem
	No	<ul style="list-style-type: none"> Insufficient information to assess whether an important risk of bias exists; or Insufficient rationale or evidence that an identified problem will introduce bias

*Minimization may be implemented without a random element, and this is considered to be equivalent to being random.

Table 5. Quality assessment tools of non-randomized studies

Evaluation criteria (Deeks et al. [21])	Downs and Black [28]	Newcastle-Ottawa [29]	Reish et al. [30]	Thomas [31]	Zaza et al. [32]
1. Background					
• Background information provided					
2. Sample					
• Retrospective/prospective			✓		
• Inclusion/exclusion criteria	✓		✓		✓
• Sample size	✓		✓		
• Representative	✓	✓	✓	✓	✓
3. Interventions					
• Clear specification of interventions	✓		✓		✓
4. Outcomes					
• Clear specification of outcomes	✓		✓		
5. Creation of groups					
• Generation of random sequence	✓		✓	✓	✓
• Concealment of allocation	✓				
• How allocation occurred	✓		✓		
• Balance groups by design		✓	✓	✓	✓
6. Blinding					
• Blind (or double-blind) administration	✓		✓		
• Blind outcome assessment	✓	✓	✓	✓	
7. Ascertainment					
• Receipt of the intervention	✓	✓	✓	✓	✓
• Attributable outcomes		✓			
8. Follow-up					
• Equal follow-up between group	✓				
• Completeness of follow-up	✓	✓	✓	✓	✓
9. Comparability					
• Baseline comparability assessed	✓		✓	✓	✓
• Prognostic factors identified	✓	✓		✓	✓
• Case-mix adjustment	✓	✓	✓	✓	✓
10. Analysis					
• Intention-to-treat analysis				✓	
• Appropriate analysis methods	✓		✓	✓	✓
11. Interpretation					
• Appropriately based on results					
• Assessed strength of evidence					
• Application/implications			✓		
12. Presentation					
• Completeness, clarity, structure	✓		✓		

관찰연구에서 사용할 수 있으나 RCT 연구 평가에는 사용할 수 없다[29]. Newcastle-Ottawa 척도는 환자-대조군 연구 평가와 코호트 연구 평가를 위하여 두 종류의 척도로 개발되어 있다. 환자-대조군 연구의 평가항목으로는 환자-대조군 선택, 환자-대조군 비교가능성, 그리고 노출 확인에 대해

평가할 수 있도록 되어 있다(Table 7A). 코호트 연구는 코호트 선택, 코호트 비교가능성, 그리고 결과 평가에 대해 평가할 수 있다(Table 7B).

평가방법은 각 항목에 대해 근거의 질이 높은 경우 '★'로 표시하도록 되어 있는데, 선택과 노출/결과에 대한 항목의 경우

Table 6. Downs and Black scale: checklist for measuring study quality [28]**Reporting:** “Yes=1,” “No=0”

1. Is the hypothesis/aim/objective of the study clearly described?
2. Are the main outcomes to be measured clearly described in the Introduction or Methods section?
3. Are the characteristics of the patients included in the study clearly described ?
4. Are the interventions of interest clearly described?
5. Are the distributions of principal confounders in each group of subjects to be compared clearly described?
“Yes=2,” “Partially=1,” “No=0”
6. Are the main findings of the study clearly described?
7. Does the study provide estimates of the random variability in the data for the main outcomes?
8. Have all important adverse events that may be a consequence of the intervention been reported?
9. Have the characteristics of patients lost to follow-up been described?
10. Have actual probability values been reported (e.g., 0.035 rather than <0.05) for the main outcomes except where the probability value is less than 0.001?

External validity: “Yes=1,” “No=0,” “Unable to determine=0”

11. Were the subjects asked to participate in the study representative of the entire population from which they were recruited?
12. Were those subjects who were prepared to participate representative of the entire population from which they were recruited?
13. Were the staff, places, and facilities where the patients were treated, representative of the treatment the majority of patients receive?

Internal validity - bias: “Yes=1,” “No=0,” “Unable to determine=0”

14. Was an attempt made to blind study subjects to the intervention they have received ?
15. Was an attempt made to blind those measuring the main outcomes of the intervention?
16. If any of the results of the study were based on “ata dredging” was this made clear?
17. In trials and cohort studies, do the analyses adjust for different lengths of follow-up of patients, or in case-control studies, is the time period between the intervention and outcome the same for cases and controls?
18. Were the statistical tests used to assess the main outcomes appropriate?
19. Was compliance with the intervention/s reliable?
20. Were the main outcome measures used accurate (valid and reliable)?

Internal validity - confounding (selection bias): “Yes=1,” “No=0,” “Unable to determine=0”

21. Were the patients in different intervention groups (trials and cohort studies) or were the cases and controls (case-control studies) recruited from the same population?
22. Were study subjects in different intervention groups (trials and cohort studies) or were the cases and controls (case-control studies) recruited over the same period of time?
23. Were study subjects randomised to intervention groups?
24. Was the randomised intervention assignment concealed from both patients and health care staff until recruitment was complete and irrevocable?
25. Was there adequate adjustment for confounding in the analyses from which the main findings were drawn?
26. Were losses of patients to follow-up taken into account?

Power

27. Did the study have sufficient power to detect a clinically important effect where the probability value for a difference being due to chance is less than 5%? Sample sizes have been calculated to detect a difference of x% and y%.

Size of smallest intervention group

A	1<n1	0
B	n1-n2	1
C	n3-n4	2
D	n5-n6	3
E	n7-n8	4
F	n8+	5

각 항목에 최대 한 개의 ‘★’을 줄 수 있으며, 비교가능성 항목에 대해서는 최대 2개의 ‘★’을 줄 수 있다. 특히 Newcastle-Ottawa 척도는 8개 항목으로 이루어진 비교적 간단한 척도로 환자-대조군 연구의 질을 평가하는 데에 유용한 척도이다.

결론

근거중심의학의 핵심 역량인 근거기반 의사결정을 위하여 근거중심 임상진료지침과 같은 기초 자료를 활용함에 있

Table 7A. Newcastle-Ottawa quality assessment scale: case-control studies [29]

NEWCASTLE - OTTAWA QUALITY ASSESSMENT SCALE CASE CONTROL STUDIES

Note: A study can be awarded a maximum of one star for each numbered item within the Selection and Exposure categories. A maximum of two stars can be given for Comparability.

Selection

- 1) Is the case definition adequate?
 - a) yes, with independent validation ★
 - b) yes, eg record linkage or based on self reports
 - c) no description
- 2) Representativeness of the cases
 - a) consecutive or obviously representative series of cases ★
 - b) potential for selection biases or not stated
- 3) Selection of controls
 - a) community controls ★
 - b) hospital controls
 - c) no description
- 4) Definition of controls
 - a) no history of disease (endpoint) ★
 - b) no description of source

Comparability

- 1) Comparability of cases and controls on the basis of the design or analysis
 - a) study controls for _____ (Select the most important factor.) ★
 - b) study controls for any additional factor ★ (This criteria could be modified to indicate specific control for a second important factor.)

Exposure

- 1) Ascertainment of exposure
 - a) secure record (eg surgical records) ★
 - b) structured interview where blind to case/control status ★
 - c) interview not blinded to case/control status
 - d) written self report or medical record only
 - e) no description
- 2) Same method of ascertainment for cases and controls
 - a) yes ★
 - b) no
- 3) Non-response rate
 - a) same rate for both groups ★
 - b) non respondents described
 - c) rate different and no designation

어서 검색된 문헌의 질을 평가하고, 평가 결과를 권고수준의 등급화 과정에 체계적으로 반영할 수 있도록 하기 위하여 국제적으로 사용되고 있는 타당한 척도들을 활용하고 그 결과를 해석하는 방법론을 습득하는 일은 진료현장에 있는 의사들에게 적·간접적으로 도움이 될 수 있다. 특히 우리나라의 경우 근거창출에 필요한 대부분의 근거가 비무작위 연구이므로 이 연구 결과들을 기반으로 의사결정을 내려야할 경우 보편타당한 근거분석 및 평가 방법론에 대한 기초연구는 무엇보다 중요하다. 무작위배정 비교임상시험 혹은 비무작위 연구 등 근거수준별로 문헌의 질 평가에 사용될 수 있는 국제적인 평가 도구들을 체계적으로 검토하고, 이중 국내에서 주로 생산되고 있는 근거의 실정에 알맞는 사용이 간편하

면서도 국제적인 기준에 부합되는 평가도구를 발굴하고 활용 방법을 제시하는 것은 근거기반 의료문화를 정착시켜 나가는 데 꼭 필요한 기초연구 중의 하나이다.

따라서 본 연구에서 근거중심 진료지침 개발과정의 핵심 구성요소로서 근거수준별로 문헌의 질 평가에 활용될 수 있는 국제적인 도구들을 검토하고 국내 현실에 적합한 도구들을 모색해 본 결과 무작위배정 비교임상시험 연구에 대한 근거의 질을 평가하기 위한 타당한 도구로는 van Tulder 척도 및 코크란 연합의 질 평가 방법이 권장될 수 있겠다. 비무작위 연구근거의 질을 평가하는 데에 유용한 도구로는 Downs and Black 도구가 있으며, 환자-대조군 연구의 질 평가를 위해 Newcastle-Ottawa 척도를 코크란 연합에서

Table 7B. Newcastle-Ottawa quality assessment scale: cohort studies [29]**NEWCASTLE - OTTAWA QUALITY ASSESSMENT SCALE COHORT STUDIES**

Note: A study can be awarded a maximum of one star for each numbered item within the Selection and Outcome categories.
A maximum of two stars can be given for Comparability

Selection

- 1) Representativeness of the exposed cohort
 - a) truly representative of the average _____ (describe) in the community ★
 - b) somewhat representative of the average _____ in the community ★
 - c) selected group of users eg nurses, volunteers
 - d) no description of the derivation of the cohort
- 2) Selection of the non exposed cohort
 - a) drawn from the same community as the exposed cohort ★
 - b) drawn from a different source
 - c) no description of the derivation of the non exposed cohort
- 3) Ascertainment of exposure
 - a) secure record (eg surgical records) ★
 - b) structured interview ★
 - c) written self report
 - d) no description
- 4) Demonstration that outcome of interest was not present at start of study
 - a) yes ★
 - b) no

Comparability

- 1) Comparability of cohorts on the basis of the design or analysis
 - a) study controls for _____ (select the most important factor) ★
 - b) study controls for any additional factor ★ (This criteria could be modified to indicate specific control for a second important factor.)

Outcome

- 1) Assessment of outcome
 - a) independent blind assessment ★
 - b) record linkage ★
 - c) self report
 - d) no description
- 2) Was follow-up long enough for outcomes to occur
 - a) yes (select an adequate follow up period for outcome of interest) ★
 - b) no
- 3) Adequacy of follow up of cohorts
 - a) complete follow up - all subjects accounted for ★
 - b) subjects lost to follow up unlikely to introduce bias - small number lost - > ____ % (select an adequate %) follow up, or description provided of those lost) ★
 - c) follow up rate < ____ % (select an adequate %) and no description of those lost
 - d) no statement

추천하고 있다. 특히 국내 문헌 혹은 근거의 대부분이 비무작위 연구임을 감안할 때 이상의 질 평가 도구들을 적절히 활용하여 비무작위 연구 중에서도 질적 수준이 높은 근거들을 구분하고, 권고 수준의 등급화 기준에 비무작위 연구들의 질적 수준을 반영하여 국내 연구 결과를 임상에서 의사결정의 중요한 근거로 활용할 수 있도록 기본 여건을 마련하는 것이 향후 중요한 과제이다.

핵심용어: 자다드 척도; 반 틀더 척도; 코크란 도구;
다운스 앤 블랙 도구; 뉴캐슬오타와 척도

REFERENCES

1. Higgins JP, Green S, editors. Cochrane handbook for systematic reviews of interventions. Ver. 5.1.0. The Cochrane Collaboration; 2011 [cited 2011 Jan 7]. Available from: <http://www.cochrane-handbook.org>.
2. Chalmers TC, Smith H Jr, Blackburn B, Silverman B, Schroeder B, Reitman D, Ambroz A. A method for assessing the quality of a randomized control trial. Control Clin Trials 1981; 2:31-49.
3. Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, Walsh S. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. Control Clin Trials 1995;16:62-73.

4. Barratt A. Evidence based medicine and shared decision making: the challenge of getting both evidence and preferences into health care. *Patient Educ Couns* 2008;73:407-412.
5. Kranke P. Evidence-based practice: how to perform and use systematic reviews for clinical decision-making. *Eur J Anaesthesiol* 2010;27:763-772.
6. Parrilla-Castellar ER, Almeyda R, Nogales E, Velez M, Ramos M, Rivera JE, Da Vila B, Torres V, Capriles J, Adamsons K. Evidence-based medicine as a tool for clinical decision-making in Puerto Rico. *P R Health Sci J* 2008;27:135-140.
7. Tiburi MF. Evidence-based medicine as viewed by key decision-makers of health plans in southern Brazil. *Health Serv Manage Res* 2008;21:185-191.
8. Stolba N, Nguyen TM, Tjoa AM. Towards sustainable decision-support system facilitating EBM. *Conf Proc IEEE Eng Med Biol Soc.* 2007;2007:4355-4358.
9. Wyer PC, Silva SA. Where is the wisdom? I. a conceptual history of evidence-based medicine. *J Eval Clin Pract* 2009;15: 891-898.
10. Collins J. Evidence-based medicine. *J Am Coll Radiol* 2007;4: 551-554.
11. Anastasiu M, Strambu V, Popa F. Evidence-based medicine, conceptual challenge or the future of daily practice? *Chirurgia (Bucur)* 2007;102:527-530.
12. Carter MJ. Evidence-based medicine: an overview of key concepts. *Ostomy Wound Manage* 2010;56:68-85.
13. Maluf-Filho F. The importance of evidence-based medicine concepts for the clinical practitioner. *Arq Gastroenterol* 2009; 46:87-89.
14. Borgerson K. Valuing evidence: bias and the evidence hierarchy of evidence-based medicine. *Perspect Biol Med* 2009; 52:218-233.
15. Isaac CA, Franceschi A. EBM: evidence to practice and practice to evidence. *J Eval Clin Pract* 2008;14:656-659.
16. Kruer MC, Steiner RD. The role of evidence-based medicine and clinical trials in rare genetic disorders. *Clin Genet* 2008;74:197-207.
17. Rogers W, Ballantyne A. Justice in health research: what is the role of evidence-based medicine? *Perspect Biol Med* 2009;52:188-202.
18. Kitto S, Petrovic A, Gruen RL, Smith JA. Evidence-based medicine training and implementation in surgery: the role of surgical cultures. *J Eval Clin Pract* 2010 Aug 4 [Epub]. DOI: 10.1111/j.1365-2753.2010.01526.x.
19. Carretier J, Bataillard A, Fervers B. The patient's role in evidence-based medicine. *J Chir (Paris)* 2009;146:537-544.
20. Soll RF. Evaluating the medical evidence for quality improvement. *Clin Perinatol* 2010;37:11-28.
21. Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakaravitch C, Song F, Petticrew M, Altman DG; International Stroke Trial Collaborative Group; European Carotid Surgery Trial Collaborative Group. Evaluating non-randomised intervention studies. *Health Technol Assess* 2003;7:iii-x, 1-173.
22. MacLehose RR, Reeves BC, Harvey IM, Sheldon TA, Russell IT, Black AM. A systematic review of comparisons of effect sizes derived from randomised and non-randomised studies. *Health Technol Assess* 2000;4:1-154.
23. Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, McQuay HJ. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials* 1996;17:1-12.
24. van Tulder M, Furlan A, Bombardier C, Bouter L; Editorial Board of the Cochrane Collaboration Back Review Group. Updated method guidelines for systematic reviews in the cochrane collaboration back review group. *Spine (Phila Pa 1976)* 2003;28:1290-1299.
25. Newell SA, Sanson-Fisher RW, Savolainen NJ. Systematic review of psychological therapies for cancer patients: overview and recommendations for future research. *J Natl Cancer Inst* 2002;94:558-584.
26. Scottish Intercollegiate Guideline Network. SIGN 50: a guideline developer's handbook 2008 [Internet]. Edinburgh: Scottish Intercollegiate Guidelines Network; 2008 [cited 2011 Jan 28]. Available from: <http://www.sign.ac.uk/pdf/sign50.pdf>.
27. National Institute for Health and Clinical Excellence. The guidelines manual 2009 [Internet]. London: National Institute for Health and Clinical Excellence; 2009 [cited 2011 Jan 27]. Available from: http://www.nice.org.uk/media/5F2/44/The_guidelines_manual_2009_-_All_chapters.pdf.
28. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *J Epidemiol Community Health* 1998;52:377-384.
29. Wells GA, Shea B, O'Connell D, Peterson J, Welch V, Losos M, Tugwell P. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses [Internet]. Ottawa: Ottawa Hospital Research Institute; c1996-2010 [cited 2011 Mar 28]. Available from: http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp.
30. Reisch JS, Tyson JE, Mize SG. Aid to the evaluation of therapeutic studies. *Pediatrics* 1989;84:815-827.
31. Thomas H. Quality assessment tool for quantitative studies: effective public health practice project. Hamilton: McMaster University; 2003.
32. Zaza S, Wright-De Agüero LK, Briss PA, Truman BI, Hopkins DP, Hennessy MH, Sosin DM, Anderson L, Carande-Kulis VG, Teutsch SM, Papaioanou M. Data collection instrument and procedure for systematic reviews in the Guide to Community Preventive Services. Task Force on Community Preventive Services. *Am J Prev Med* 2000;18(1 Suppl):44-74.



Peer Reviewers' Commentary

근거중심 의학(Evidence Based Medicine)은 특정질병의 원인, 진단, 치료, 예방에 대하여 연구한 연구문헌들을 체계적으로 고찰하여 과학적으로 유효하고 합리적인 결과를 진료에 활용하는 것이다. 최근 각 학회를 중심으로 만들어 지고 있는 임상진료지침도 근거중심의학의 기본으로 하여 작성되고 있다. 이때 검색된 연구문헌에 대한 근거의 질 평가는 매우 중요하다. 근거문헌에 대한 질 평가를 통해 연구설계, 수행 및 분석에서 나타날 수 있는 비뚤림(bias)의 가능성을 없애거나 최소화할 수 있다. 논문의 결론을 수용할 것인지를 결정하고, 추가 연구가 필요한지에 대하여 판단하기 위한 수단으로 근거문헌에 대한 질 평가는 필수적이다. 본 논문은 근거중심의학의 위한 기초 자료를 활용함에 있어서 검색된 문헌의 질을 평가하고, 평가결과를 권고수준 등급화 과정에 체계적으로 반영할 수 있는 국제적 도구들을 검토하여 국내 현실에 적합한 평가 방법을 제시하고 있다. 근거중심의학의 방법론을 표준화 하는데 도움이 될 것으로 판단되며 앞으로 더욱 많은 연구로 우리나라의 근거중심의학이 발전할 것을 기대해 본다.

[정리:편집위원회]