

체계적 고찰 연구의 개요

안 형 식 · 김 현 정 | 고려대학교 의과대학 예방의학교실

An introduction to systematic review

Hyeong Sik Ahn, MD · Hyun Jung Kim, PhD

Department of Preventive Medicine, Korea University College of Medicine, Seoul, Korea

Systematic review had become one of the important research area in medicine. Systematic review can be demonstrating benefit or harm of an intervention when results of individual studies are inconclusive. While narrative reviews can often include an element of selection bias, systematic reviews typically involve a comprehensive plan and search strategy with the goal of reducing bias by identifying, appraising, and synthesizing all relevant studies on a particular topic and investigation of heterogeneity among included studies. Systematic reviews typically include a meta-analysis component which involves using statistical techniques to synthesize the data from several studies into a single quantitative estimate or summary effect size. Systematic review overcomes the limitation of small sample sizes by pooling results from a number of individual studies to generate a single best estimate. Although systematic reviews are published in academic forums, the Cochrane Collaboration is a widely recognized international and not-for-profit organization that promotes, supports, and disseminates systematic reviews and meta-analyses on the efficacy of interventions in the health care field. Systematic review has become a popular and powerful tool. If rigorously conducted, it is essential for evidence-based decision making in clinical practice as well as on the health policy level.

Key Words: Systematic Review; Meta-analysis; Evidence-based medicine; Medical literature search; Publication bias

서론

의학연구에서는 동일한 주제임에도 불구하고 연구에 따라 상반되는 결론을 제시할 수도 있고, 혹은 너무 정보가 많아 어느 것이 올바른 의사결정인지에 대한 판단이 힘든 경우가 발생한다. 체계적 고찰(systematic review)은 이처럼 개인

이 수행하기 어려울 정도로 많은 정보 속에서 근거를 종합하여 신뢰할 수 있는 객관적 근거를 제공하는 것이다. 고찰 연구는 과거에는 독창성과 중요성이 떨어지는 영역으로 여겨졌으나, 일차 연구만으로는 어떤 문제에 결론을 내리기 어려운 상황에서 문헌을 포괄적으로 검토해서 결론을 내리는 것이며, 근거중심의학의 등장과 함께 의학 분야의 중요한 연구 분야로 자리 잡았다. 이 글에서는 체계적 고찰의 수행과 해석 및 향후 방향에 대해 소개하고자 한다.

Received: October 20, 2013 Accepted: November 2, 2013

Corresponding author: Hyeong Sik Ahn
E-mail: ahnhs@korea.ac.kr

© Korean Medical Association

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

역사적 전개

의학 분야에서 체계적 고찰을 처음 제시한 사례는 1904년

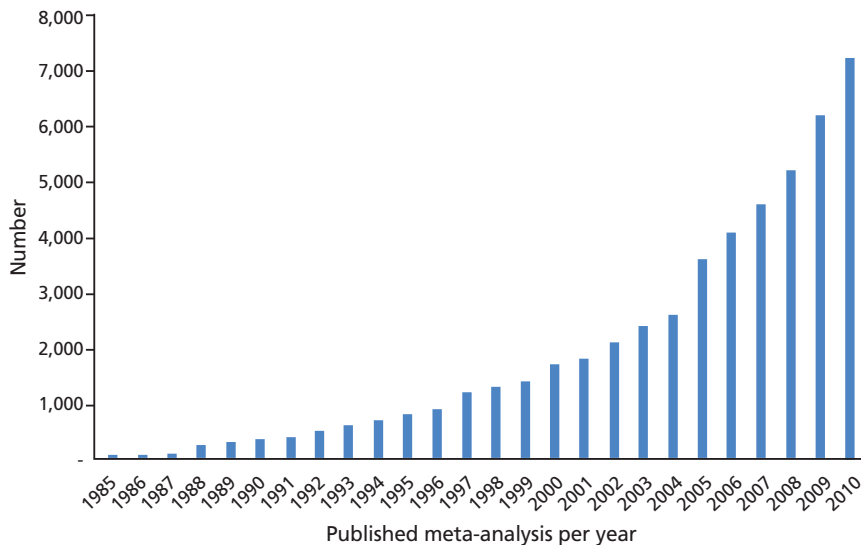


Figure 1. Number of systematic reviews (meta-analysis) MEDLINE from 1985 to 2010.

에 Karl Pearson이 장염 예방에 혈청주입법의 자료를 통합한 사례를 들 수 있다. "자료의 수가 너무 적어서 명확한 결론을 내리기 어렵다"는 그의 문제제기는 지금도 유효하다고 할 수 있다[1]. 체계적 고찰의 본격적인 등장은 1970년대 이후에 들어서 이루어졌다. 1976년에 Glass [2]가 '메타분석'이라는 용어를 사용하기 시작하였고, Mulrow [3]는 서술적 고찰의 오류 가능성을 지적하고 체계적 고찰의 중요성을 실증적으로 제시하였다.

체계적 고찰을 현재의 모습으로 갖추는 데에 Archie Cochrane의 업적이 크다. 그는 스코틀랜드 태생의 의학자로서 의료서비스의 효과에 대한 문제제기와 아울러 근거중심의료의 필요성을 강조하고 실천하였다. 1972년도에 '효과와 효율: 의료서비스의 효과에 대한 비판적 고찰(Effectiveness and efficacy: random reflections in health services)' [4]라는 저서를 통하여 의료서비스의 효과를 입증하기 위한 무작위 임상시험의 중요성을 피력하였다. 1979년에 그가 쓴 에세이에서 '의학의 각 분야별로 무작위 임상시험의 결과를 수집하고 엄격히 평가하여 그 결과를 주기적으로 개정하여 보급하는 체계가 없는 것은 우리 전문직의 수치이다'[5]라는 언급은 코크란 체계적 고찰의 탄생의 배경이 된 유명한 말이다. 이후 영국 옥스퍼드에서 산부인과 영역에서 무작위 임상시험의 자료를 통합하기 시작하여 1974년에서 1985년까지 3,500여 개의 문헌을 요약하여 600개

의 체계적 고찰을 완성하였다. 사후인 1992년에 영국에 그를 기리어 코크란 센터가 설립되었고 이는 이후 체계적 고찰의 발전에 크게 기여하였다[6].

체계적 고찰의 코크란 및 기타 체계적 고찰, 메타분석 등의 연구를 메드라인(MEDLINE)을 통하여 검색해보면 전 세계적으로 급증하고 있는 추세임을 알 수 있다. 1990년까지 연간 수백 개에 불과하던 것이 2010년경에는 연간 8,000개에 이르도록 급격하게 증가하였다(Figure 1). 이는 세계적으로 출간되는 전체 논문 중 약 1.5%에 달하는 비

중으로, 체계적 고찰이 의학의 중요한 연구영역으로 성장하였음을 알 수 있다. 또한 코크란 연합의 산물인 코크란 리뷰는 100여 개국의 28,000명의 저자가 참여하여 5,000개 이상의 리뷰가 수록되어 있다. 우리나라에서도 최근 들어 체계적 고찰이 근래 몇 년에 들어 많이 증가하여 최근 3년간 국내에서 이루어진 체계적 고찰은 275 여개에 달하고 있다[7].

체계적 고찰의 종류와 자료원

1. 체계적 고찰의 종류

체계적 고찰은 기존 문헌의 연구 결과들을 정리한다는 의미에서 고찰이라 할 수 있지만, 기존의 서술적 고찰(narrative review)과는 다르다. 서술적 고찰은 전문가의 식견을 바탕으로 기존의 문헌들을 고찰하고 저자의 주관적인 관점에서 종합하는 것이다. 그러나 방법이 체계적이지 못하며 따라서 저자의 주관에 따라 다른 결론이 내려질 수 있다. 즉 기존의 서술형 고찰은 광범위한 영역과 주제를 다루고 문헌의 선정과 종합 등의 방법이 체계적이지 않아 주관적 오류의 가능성이 있다고 하겠다. 반면 체계적 고찰은 기존 연구 자료를 사용하되 과학적이고 객관적인 방법에 따라 특정하게 좁혀진 주제에 대하여 엄격한 방법론을 사용하여, 종합된 추정치

Table 1. Comparison of narrative and systematic reviews

Feature	Systematic review	Narrative review
Question	Clearly defined topic or question	Overview of an area
Search strategy	Explicit search protocols	Systematic search protocol not used
Selection criteria	Using a protocol that specifies inclusion, exclusion criteria	Do not use to an explicit, predetermined protocol
Methodological appraisal of primary studies	Include objective appraisal of individual studies	Do not always use level of evidence rating system
Synthesis of results	Quantitative often includes meta-analysis	Unknown/qualitative

등을 제시하는 것이다[8]. 서술적 고찰과 체계적 고찰의 구체적인 차이는 Table 1과 같다.

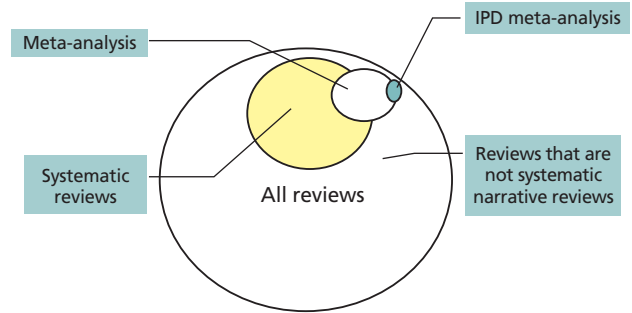
한편 메타분석(meta-analysis)은 두 개 이상의 일차연구를 통계적으로 결합하여 통합 추정치를 제시하는 것이다. 즉 통계적으로 문헌을 통합한 연구인 것이다. 대부분의 메타분석은 체계적 고찰의 일부로서 수행된다. 그러나 통계학적인 추정치결합이 필요 없는 경우에는 메타분석 없이 체계적 고찰이 수행될 수 있으며, 이 경우의 체계적 고찰은 질적인 평가만을 수행하여 결과를 제시하기도 한다[9].

Individual patient data (IPD) 메타분석은 분석할 자료가 출판된 논문의 저자로부터 직접 얻은 개별 연구 대상자의 자료인 경우이다. IPD 메타분석이 기존 문헌에 대한 메타분석보다 더 정확한 결과를 도출할 수 있으나 자료획득의 어려움으로 인해 활발히 수행되지는 않는다.

고찰, 체계적 고찰, 메타분석, IPD 메타분석의 관계는 Figure 2와 같다[10]. 체계적 고찰의 주제는 전통적으로는 치료효과를 다루지만, 진단, 예후, 발생률, 및 질적 연구 등 다양한 주제에 대해 점점 그 영역이 넓어지고 있다.

2. 체계적 고찰의 장점

체계적 고찰의 장점은 기존 일차연구들의 결과를 종합하여 근거를 요약한 결과를 제시하여 비필립이 최소화된 정보를 제공할 수 있다는 점이다. 체계적 고찰과 흔히 같이 수행되는 메타분석의 큰 장점은 표본 수를 증가시키는 데 있다. 작은 규모의 연구를 취합하면 제2형 오류(type II error)의 위험성을 줄여 결론의 신뢰도를 높인다.

**Figure 2.** Relationship of review studies; narrative review, systematic review, meta-analysis, and individual patient data (IPD) meta-analysis.

이처럼 검정력을 증가시켜 후속연구가 수행될 필요없이 효과를 검증함으로써 효과적인 치료법을 보다 빠르게 도입하는데 기여할 수 있다. 아울러 체계적 고찰은 일반화 가능성을 높일 수 있다. 개별연구는 연구대상자가 제한되지만, 체계적 고찰은 다양한 인구집단에 대한 치료법의 효과를 파악할 수 있기 때문에 대상자를 확대하여 일반화 할 수 있기 때문이다. 또한 다양한 인구집단을 대상으로 한 연구들을 분석하여 집단별로 효과의 차이를 검증함으로써 어느 집단에서 효과가 있는지를 제시할 수도 있다.

3. 체계적 고찰의 자료원

체계적 고찰을 찾을 수 있는 자료원은 몇 가지가 있는데, 메드라인과 코크란 라이브러리(Cochrane library)가 가장 대표적 데이터베이스이다. 메드라인에서는 “clinical queries” 기능을 사용하여 체계적 고찰이나 메타분석 논문을 찾을 수 있고, 혹은 “type of article”에서 “meta-analysis”로 제한을 하여 문헌을 검색할 수도 있다. 코크란 연합에서 제공되는 코크란 라이브러리(www.thecochranelibrary.com)는 코크란 리뷰와 York Database of Abstracts of Reviews of Effectiveness (DARE) 등 다른 체계적 고찰들이 수록되어있는 데이터베이스이다. 코크란 라이브러리에서는 “advanced search”를 사용하여 코크란 자료와 DARE의 자료를 구분하여 체계적 고찰 및 메타분석을 검색할 수 있다. 코크란 라이브러리는 유료이지만, 현재에 대부분의 우리나라 의과대학에서 구독하고 있어 이를 통해 접근할 수 있다.

이외 체계적 고찰이나 메타분석만 전문적으로 검색할 수 있는 검색엔진으로는 TRIP (<http://www.tripdatabase.com>), Sumsearch (<http://sumsearch.uthscsa.edu>) 등이 있다. 이들은 이차검색을 통해 검색결과를 제시하며 논문만이 아니라 기관의 체계적 고찰 보고서 등도 자료원 별로 제시하고 있다.

체계적 고찰의 수행과정

체계적 고찰 수행과정은 주제선정과 핵심질문 도출, 문헌 검색 및 선택, 문헌에 대한 비판적 평가, 메타분석, 결과평가 및 출간과 같이 다섯 단계로 나눌 수 있다.

1. 주제선정과 핵심질문

체계적 고찰의 주제선정은 의학적 의미가 있고 독창적이어야 하며, 주어진 질문에 답을 줄 수 있는 가능성이 있어야 한다. 체계적 고찰 연구의 가치는 어떤 주제를 선정하느냐에 따라 상당부분 좌우되기에 주제 선정에는 그 분야의 임상적 중요성, 연구 경향 등에 대한 면밀한 검토가 필요하다. 또 하나 중요한 점은 사전적(a priori) 연구기획이 중요하다는 점이다. 사후에 결과를 보고 분석하는 것은 여러 오류의 가능성이 있기 때문이다. 따라서 사전에 가능한 한 상세한 연구계획을 세우는 것이 바람직하다. 현실적으로는 리뷰할 수 있는 논문의 종류와 개수가 체계적 고찰의 가능성을 제한하기에 이를 미리 검토하여 시작하여야 한다.

연구주제가 결정되면 핵심질문을 만들어 주제를 명확히 정의하고 구체화 할 필요가 있다. 핵심질문은 네 가지 요소로 구성되며, 첫째는 관심 환자집단을 어떻게 정의할 것인가(patient), 둘째는 어떠한 중재법(intervention)에 대한 결과를 평가하고자 하는가, 셋째는 무엇을 기준으로 평가할 것인가 즉 무엇과 비교하여 결론을 내릴 것인가(comparator), 넷째는 결과(outcome) 측정방법은 어떻게 정할 것인가 등이다. 이와 같이 관심질문을 구체적으로 정의하는 것을 'PICO'의 설정이라고도 한다. 체계적 고찰에서는 연구설계가 중요하기에 어떤 연구설계(study design)를 찾을 것

인지를 핵심질문에 포함하여 'PICOS'로 표현하기도 한다.

2. 문헌검색 및 선택

핵심질문을 형성한 후에는 이에 맞는 검색전략을 수립하여 문헌을 찾아야 한다. PICO의 선정을 통해 구체화된 임상질문은 다시 각 요소에서의 주요 개념어를 선정하여 검색을 수행한다. 문헌검색의 목적은 답을 구할 수 있는 체계적 고찰의 특성에 맞는 근거를 찾아내는 것이며, 이는 최신의 근거를 담고 있는 자료원을 선택하고 검색하는 일련의 과정으로 구성된다.

1) 자료원의 종류

(1) 일반 검색원

메드라인은 가장 우선적인 데이터베이스이다. 메드라인에는 약 천만 개에 이르는 참고문헌 등 거대한 규모와 광범위한 분야를 망라하고 있기 때문에 체계적 고찰의 문헌을 위해서는 일정한 숙련이 필요 하다. 메드라인과 유사한 데이터베이스인 EMBASE는 네덜란드 Elsevier사에서 제작한 검색엔진으로 생물의학 및 약학 관련 정보를 제공한다. 현재 4,800여종의 저널이 색인되어 있으며 이중 메드라인에 색인되어 있지 않은 저널은 1,800종인 것으로 보고되고 있으며 이외에도 Science Citation Index, Current Contents Connect 등이 있다. 체계적 고찰에서는 적어도 메드라인과 EMBASE는 포함할 것을 추천한다.

(2) 특정주제별 검색원

의학의 여러 분야 중 특정 주제에 적합한 근거를 위주로 제공하는 검색원이나 혹은 검색엔진이 있다. 예를 들어 PsyINFO는 미국심리학회에서 개발된 검색엔진으로 전 세계 심리학, 정신의학 및 관련 문헌에 대한 초록을 제공하고 있다. 이와 같이 특정 분야에 대한 전반적인 자료를 수록한 검색엔진은 CINAHL, NHS Economic Evaluation Database (NHS EED), MIDIRS, SLGIE 등이 있다.

(3) 국내 데이터베이스

국내의 주요 검색데이터베이스는 KoreaMed, 국립중앙도서관, 국회도서관, 한국교육학술정보원, 한국과학기술정보연구원 과학기술학회마을, 국내의학학술지초록 검색(KMbase), KSI 한국학술정보가 있다. 그러나 국내검색의

경우 검색범위가 서로 중복되어 있고 검색어가 아직 개념화되어 있지 않아 검색을 위해서는 노력이 요구되고 있다.

2) 검색전략 수립

주제 영역에 대한 주요 개념어의 선정이 이루어졌다면, 각 PICO 요소 검색어를 적절한 조합할 필요가 있으며 각 개념어들 간에 연산자를 이용하여 정확한 질문을 만드는 과정을 통해 검색을 수행하여야 한다. 데이터베이스를 선정하였다면 연구목적에 적합한 민감도(sensitivity)와 특이도(specificity)를 가지는 검색전략을 수립하여야 한다. 민감도란 검색전략의 포괄성을 의미하며 정확도란 부적절한 문헌을 배제하는 검색능력을 의미한다. 체계적 고찰을 수행하기 위해서는 기본적인 모든 문헌을 포함하기 위한 포괄적인 검색전략을 수립하도록 권장한다. 즉 민감도가 높은 전략을 구사하여야 하는 것이다. 예를 들어, 코크란연합에서는 민감도가 높은 전략인 Highly Sensitive Search Strategy (HSSS)를 미리 만들어 제공하며 이를 활용할 수 있다. 그러나 핵심 질문과 연관성이 낮은 문헌들을 많이 포함하게 되기에 많은 노력을 필요로 하는 경우가 많다.

3. 문헌평가

검색된 논문은 적절한 평가과정을 거쳐야 한다. 방법론적인 타당성이 결여되어 있다면 체계적 고찰의 질문에 대해 적절한 해답을 줄 수 없고, 반면 정교하게 설계된 연구들은 진실에 가까운 결과를 줄 수 있기 때문이다. 따라서 문헌의 질을 평가하는 것은 체계적 고찰 결과에 영향을 주기 마련이며 질평가에 따라 포함문헌을 제한하거나 질적인 수준별로 별도 분석하는 경우도 많다.

연구결과 분석단계에서 문헌평가의 주요내용은 연구설계, 수행, 분석상의 오류로 인해 실제로 발생한 사실을 연구 내에서 추론하지 못하게 되어 발생한 문제가 있는지를 검토하는 것이다.

1) 무작위 임상시험 문헌평가 도구

무작위배정 비교임상시험은 선택비뚤림, 실행비뚤림(중재 시행여부에 대한 정보의 충실성, 눈가림 등), 결과확인비뚤림(결과평가의 정확성, 눈가림 등), 탈락비뚤림, 보고비뚤림(출판비뚤림과 결과의 선택적 보고) 등의 영역을 평가한다.

무작위 임상시험 평가도구가 있으나 코크란 연합에서 제시한 위험은 평가도구인 Risk of Bias Tool이 대표적이며 이는 다음과 같은 항목으로 이루어져있다[11].

(1) 순서생성

환자를 무작위 방법을 사용하여 치료군과 대조군에 배정하였는지에 관한 것이다.

(2) 배정 은폐

무작위 배정 과정이 임상시험을 계획한 의사에게 비밀로 이루어졌는지에 관해서 평가한다.

(3) 참여자, 연구자, 결과평가자 눈가림법

환자 및 연구자, 연구결과 측정에서 맹검법이 수행되었는지를 평가한다.

(4) 불완전한 결과자료

참여한 모든 환자들의 임상 양상이 충분한 기간동안 추적 관찰되었는지 여부를 평가한다.

(5) 선택적 결과보고

선택적 결과보고는 애초에 기획된 변수 중 일부만 선택하여 보고하는 것이다. 의미 없는 결과를 배제하고는 의미 있는 결과 변수만을 보고하였는지 등을 평가한다.

(6) 타당도를 위협하는 다른 잠재적 비뚤림

조기 종료, 기저상태 차이, 비눈가림법 연구에서 블록 무작위 배정 등을 평가한다.

2) 비무작위 임상시험 문헌평가 도구

비무작위 연구의 비뚤림 평가영역도 무작위배정 비교임상시험과 유사하게 선택비뚤림, 실행비뚤림, 결과확인비뚤림, 탈락비뚤림, 보고비뚤림 등을 평가한다. 다만 무작위 배정, 배정 눈가림 등은 포함되지 않으며, 비무작위 임상시험에서 평가하여야 하는 중요한 것은 교란변수이다. 코크란 그룹에서는 교란변수 평가의 결과보고는 교란변수에 대해 1) 모든 군에서 교란변수에 대해서 동일한 값을 갖는지, 2) 교란변수에 대해서 두 군 사이에 균형이 있는지, 3) 교란변수에 대해서 짝짓기를 했는지, 4) 교란변수에 대해서 보정을 했는지 등을 평가하도록 권장하고 있다.

비무작위 연구는 연구설계가 다양하기 때문에 연구설계에 따라 구분하여 평가하기도 한다. 예를 들어 전향적 연구와 후향적 연구, 코호트 연구 그리고 환자대조군 연구와 구분하

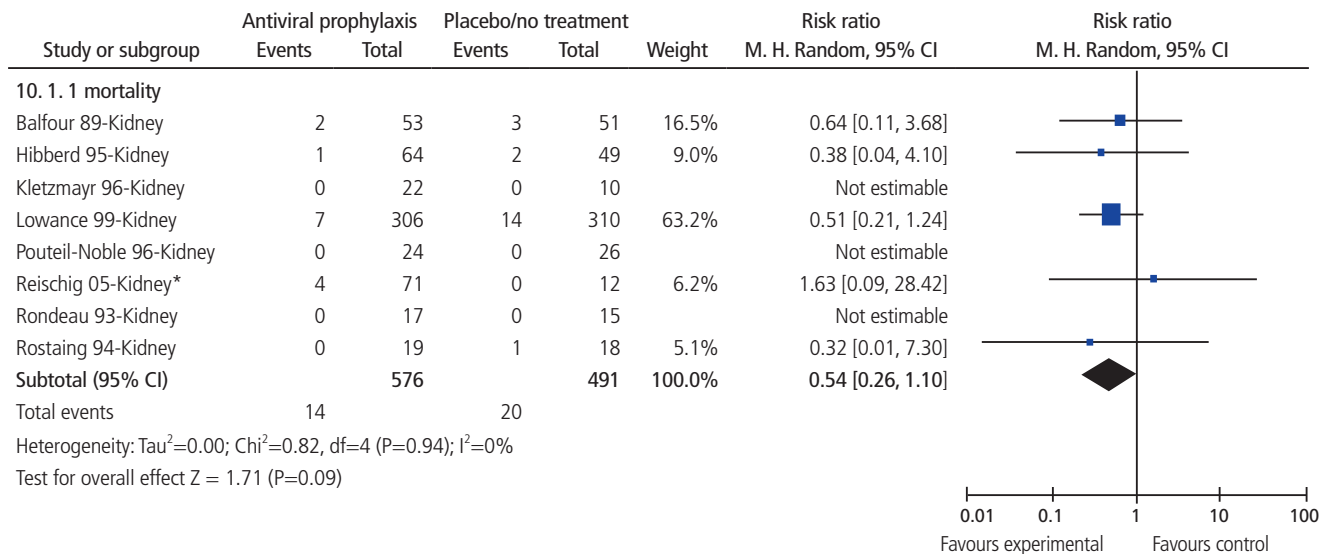


Figure 3. Example of forest plot; effect of antiviral prophylaxis after kidney transplantation. CI, confidence interval.

여 평가하여 그 결과를 제시할 수도 있다. 뉴캐슬 오타와 척도(New Castle Ottawa Scale)는 연구설계 별로 평가 항목을 제시하고 있다.

4. 메타분석

메타분석은 여러 개의 연구결과를 요약하여 종합적인 결론을 이루기 위한 방법으로 통계적 기법을 이용하여 연구들의 공통적 효과의 결과를 종합하는 것이다. 메타분석에서는 여러 연구를 종합할 때 연구의 특성별로 상이한 가중치를 두게 되는데, 기본원칙은 더 정확한 결과를 내는 연구 즉 표본수가 많거나 추정치가 클수록 평균을 계산할 때 더 많은 가중치를 준다.

메타분석과정은 자료의 내용을 표로 요약하고 그래프로 정리하는 방식으로 진행된다. 자료 요약은 각 연구 자료를 출판년도, 설계, 대상 피험자, 시험적 중재, 결과변수, 연구의 질 평가 등 전반적인 파악이 가능하도록 요약한다. 그래프는 연구결과들은 시각적으로 제시하는 방법으로, 메타분석에서는 숲그림(forest plot)이라고 하는 그래프로 표현된다. 이는 각각의 연구로부터 얻어진 결과의 점추정치와 신뢰구간을 제시하고 종합적인 요약추정치를 평행하게 그려 나감으로써 나무들이 뺨뺨하게 서 있는 숲과 같은 형태로 표시하는 것이다. 이와 같은 제시방법은 종합적인 치료효과와 크기에 대한 가늠뿐만 아니라 개별 연구결과들간의 이질

성 정도를 시각적으로 평가할 수 있다(Figure 3). 이런 결과를 통해 메타분석의 결과는 치료효과의 추정치 및 신뢰구간, 통계적 유의성 검정결과, 연구간 치료효과의 이질성 검정을 제시하며, 숲그림에서는 이를 시각적으로 표현하여 제시한다.

메타분석은 소프트웨어를 이용하여 수행하게 되는데 몇 가지 주요한 소프트웨어가 있다. 코크란 연합에서 제공하는 Revman이라는 프로그램은 코크란 홈페이지를 통해 무료로 다운로드 받을 수 있는데, 메타회귀분석이나 베이esian분석 등을 제외한 대부분의 메타분석을 다룰 수 있고, 숲그림이나 깔때기그림 제공하며 코크란 리뷰를 할 때 유용하다. 상용프로그램으로서 Comprehensive Meta-analysis나 R 등의 프로그램이 있다. 이들 프로그램은 메타분석 전용이기에 비교적 사용하기 쉽게 다양한 방식의 기본적인 메타분석과 그림 그리기 메뉴를 제공하여 메타회귀분석 또한 가능하다.

SAS나 STATA 등 일반적인 통계소프트웨어에서도 메타분석을 할 수 있다. 매크로를 이용하여 메타분석을 할 수 있으며 이질성 검정, 메타회귀분석 등이 가능하다. 진단분석 등 복잡한 자료를 운영하는데 적합하지만, 초보자의 경우 통계 전문가의 자문을 받을 필요가 있다. 아울러 1990년 이후 체계적 고찰과 메타분석을 위한 다수의 책들이 출간되어 참고할 수 있으며, 특히 코크란 핸드북은 PDF파일의 형태로 다운로드가 가능하다[12]~[16].

체계적 고찰의 한계: 일차 연구의 질, 이질성 검정 및 출판비뒀림

체계적 고찰의 결과는 주의 깊게 해석되어야 한다. 임상적 의미에 대한 검토 없이 무작정 통계결과를 받아들이는 것은 결론 오류를 초래할 수 있으며, 메타분석의 기계적인 해석은 진실과 다른 연구 결과를 제시하게 된다. 역사적으로 기존 소규모 연구를 대상으로 한 메타분석의 결과가 이후 대규모 임상시험으로 뒤집어진 결과가 도출된 예들이 있다. 메타분석의 해석은 다음과 같은 측면에서 검토되어야 한다.

1. 포함된 일차 연구의 질적 수준

아무리 방법론적으로 잘 수행된 체계적 고찰 또는 메타분석이라 하더라도 기본적으로 포함된 연구의 질적 수준이 낮다면, 양질의 결과를 도출할 수는 없다. 즉 메타분석의 추론 수준은 포함된 일차연구의 질적수준에 좌우되기 마련이다(garbage in garbage out). 하나의 체계 질적수준의 높은 연구와 낮은 연구가 같이 포함된 경우라면 민감도 분석을 수행하여 제시하여야 한다.

2. 연구의 이질성

체계적 고찰은 포함된 일차연구가 비뒀림 없는 추정치(unbiased estimate)를 제시하고 있음을 전제로 한 것이다. 즉 포함된 연구는 동질하다는 전제에서 출발하며, 이런 전제가 맞다면 각 연구결과간의 차이는 우연히 일어난 것일 뿐 실제 차이가 존재하는 것은 아니다. 그러나 포함된 연구가 서로 상이하여 이질성이 존재한다면 이는 전제를 위반하는 것으로 공통적인 치료효과를 종합하는 것이 적절하지 못할 수 있다. 이처럼 체계적 고찰에서는 포함된 연구의 이질성 여부가 연구결과의 타당성에 큰 영향을 주는 중요 요소이다.

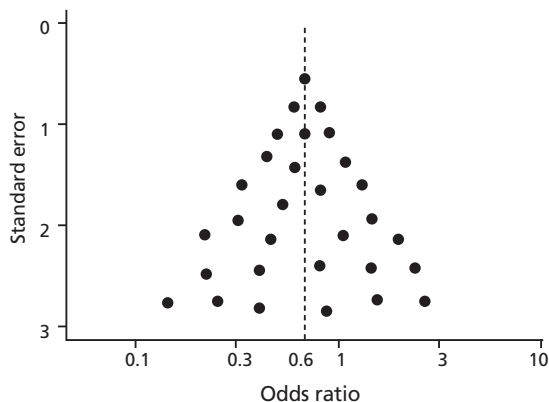
이질성 여부는 메타분석 그림을 통해 시각적으로 파악이 가능하거나 혹은 통계적 검정을 통해서 확인할 수도 있다. 시각적으로 확인하는 방법은 숲그림의 그래프를 통해 연구 간 신뢰구간 및 치료효과 추정치에 공통적인 부분이 있는지 확인하는 것이다. 통계적 검정은 수치를 통해 검정하는 것인데 이 중 대표적인 것은 I^2 로서 연구들 간의 이질성을 측

정하는 수치로 제시한다.. 숲그림을 통해서 95%의 신뢰구간이 겹치지 않거나 I^2 가 큰 경우 이질성이 높다고 판단할 수 있으며 이 경우에는 연구들 간 이질성의 원인을 찾아야 한다[17].

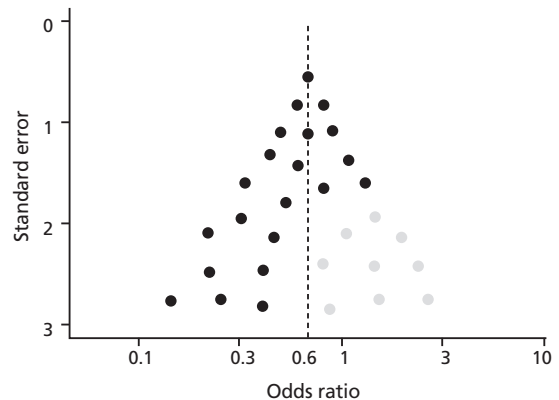
이질성의 원인은 연구대상의 상이성, 임상적 중재방식의 차이, 병용요법 사용 여부, 결과변수 측정방식의 차이, 연구 설계상의 차이, 연구의 질 등에서 비롯될 수 있으며 파악되지 않은 다른 요인들이 있을 수 있다. 이질성의 원인이 파악되면 이를 자료해석에 적극 반영할 수 있다. 예를 들어 Colditz 등[18]이 분석한 결핵예방 BCG 접종효과에 관한 메타분석에서 추운지역과 따뜻한 지역 간의 예방효과의 뚜렷한 차이를 보였고 이 경우 통합해서 분석하는 것보다는 지역별로 나누어 예방접종의 효과를 달리 해석 하는 것이 합리적이라고 할 수 있다. 이질성 원인을 파악하기 위한 통계적 분석 방법은 세부분 분석이나 민감도 분석, 동질성(sensitivity analysis), 메타회귀분석(meta-regression) 등이 있다. 관찰된 이질성이 충분히 설명되지 못한다면, 제시된 연구의 타당성에 대해서는 결과를 의심할 필요가 있고 주의 깊게 해석되어야 한다.

3. 보고비뒀림 및 출판비뒀림

비뒀림에 대한 논의는 대부분 개별 연구단위에서 이루어지는 것이다. 그러나 체계적 고찰에서 가장 중요한 비뒀림은 개별연구단위 보다는 수행된 전체 연구 중에서 어느 정도가 보고되느냐 하는 출판물의 존재여부에 관한 것이다. 이를 보고비뒀림(reporting bias)이라고 한다. 보고비뒀림 중 하나인 출판비뒀림(publication bias)은 연구의 결과에 따라 출판여부가 결정되어, 출판된 연구만으로 종합할 때 결과가 왜곡되는 것이다. 이러한 예는 흔한 것으로 추정되는데 수행된 연구 중에서 50%만이 데이터베이스에 검색될 수 있는 형태로 출판된다고 한다[19]. 문제는 출판되지 못하는 것이 우연히 일어나는 것이 아니라 연구결과에 따라 선택적으로 일어난다는 점이다. 통계적으로 유의한 결과를 보인 자료만이 출판되었고, 통계적인 유의성을 보이지 못했던 연구결과들이 배제된 상태에서 분석이 이루어지게 되는 경우, 실제보다 훨씬 더 낙관적으로 부풀려진 결과를 제시하게 되어 의로서비



A



B

Figure 4. Example of funnel plot (hypothetical). (A) symmetrical plot in absence of bias. (B) asymmetrical plot in presence of reporting bias.

스에 나쁜 영향을 주게된다.

Turner 등[20]은 미국 Food and Drug Administration (FDA)에 제출된 12개 항우울제 약의 의사진료에 임상시험 보고서에 대해 실제의학저널에 어느 정도 출판되는지를 조사하였다. 출판된 임상시험은 51개였는데 이 중 94%가 임상시험이 효과가 있었던 반면, FDA에 제출된 74개 임상시험을 모두 포함하였을 때는 51%만이 효과가 있었다. 이들 자료의 메타분석에서는 출판된 임상시험만으로는 69%의 효과를 보인 반면 FDA에 제출된 임상시험 전체를 포함한 경우는 11%의 효과만이 있었다.

또 하나의 예로는 심근경색 이후 정맥 마그네슘 투여의 효과에 관한 체계적 고찰을 들 수 있다. 메타분석 연구에서는 효과가 있었으나[21], 이후 시행된 대규모 임상시험인 심근경색증 생존자 임상시험(Fourth International Study of Infarct Survival)에서 결과가 뒤집어진다[22]. 애초의 메타분석의 결과는 효과가 없는 시험이 출판되지 않았기 때문인 것으로 판단되었다.

출판비뚤림을 줄일 수 있는 가장 좋은 방법은 보다 완벽한 문헌검색을 하는 것이다. 이를 위해서는 다양한 데이터베이스를 찾고, 컨퍼런스의 초록을 포함하고, 임상시험 등록자료를 찾거나 저자나 제약회사를 접촉하여 출판되지 않은 임상 시험을 포함하는 것이다. PRISMA (Preferred Reporting Items for Systematic Reviews)는 체계적 고찰의 보고양식인데 문헌의 검색일자, 검색전략을 기술하여 문헌검색이 재현될 수 있도록 권고하고 있다[23,24]. 컨퍼런

스의 초록을 포함할 것인지는 논란의 소지가 있지만 주의 깊은 해석을 전제로 가능하다.

출판비뚤림의 개입 여부를 일차적으로 탐색하는 방법은 깔때기그림을 통해 시각적으로 판단하는 것이다[25]. 깔때기그림은 그림의 형태가 깔때기를 뒤집어 놓은 것처럼 보여야 한다고 해서 붙여진 이름으로, 가로축에는 각 연구로부터 얻어진 교차비나 상대위험도와 같은 치료효과의 추정치를, 세로축에는 제공된 추정치의 정밀도를 나타내는 표준오차의 역수 또는 연구대상수 등을 설정하여 그린 산점도이다(Figure 4). 깔때기그림에 공백이 관찰될 때에는 출판되지 않는 연구가 존재하여 출판비뚤림이 있다고 추정할 수 있는데 이 공백을 채워넣었을 경우를 가정하며 출판비뚤림에 의한 결과를 보정하는 방법론적인 시도(trim and fill) 등이 있으나 실제 적용에는 주의가 필요하다.

연구가 출판되지 않는 것은 저널에서의 게재 거부 등 여러 요인이 있으나 가장 중요한 것은 스폰서의 의도적인 누락이며, 의도적이기에 교정되기가 더욱 어려운 문제이다. 출판비뚤림은 체계적 고찰의 타당성을 제해하는 가장 큰 요인이라고 할 수 있으며 앞으로 사회적으로도 많은 노력이 필요한 영역이다.

4. 체계적 고찰에 대한 평가

체계적 고찰은 단 하나의 연구가 가지는 근거보다 신뢰성이 높다고 할 수 있으나 체계적 고찰의 질적 수준이 다양하므로 고찰은 면밀히 평가되어야 한다. 따라서 먼저 체계적

Table 2. Appraisal tool of systematic review; AMSTAR (A Measurement Tool to Assess Systematic Reviews)

Feature
Was an 'a priori' design provided?
Was there duplicate study selection and data extraction?
Was a comprehensive literature search performed?
Was the status of publication (i.e., grey literature) used as an inclusion criterion?
Was a list of studies (included and excluded) provided?
Were the characteristics of the included studies provided?
Was the scientific quality of the included studies assessed and documented?
Was the scientific quality of the included studies used appropriately in formulating conclusions?
Was the scientific quality of the included studies used appropriately in formulating conclusions?
Was the likelihood of publication bias assessed?
Was the conflict of interest stated?

고찰의 질을 평가한 후 자신의 환자에게 적용할 수 있는지를 검토하고 임상적 결정을 해야 한다. 질평가를 할 수 있는 도구 중 A measurement Tool to Assess the Methodological Quality or Systemic Review (AMSTAR)는 체계적 고찰의 방법론적 질을 평가하는 도구 중 가장 최근에 개발된 것으로 타당성 검토가 이루어져 유용하다[26]. AMSTAR에는 11개 항목이 있으며 Table 2에 수록되었다.

우리나라에서의 과제

우리나라에서 체계적 고찰에 관한 활동은 근래에 많이 개선되었다. 국내 체계적 고찰의 연구현황 분석에 따르면 2008–2010년의 3년간 총 220건의 체계적 고찰 연구가 이루어졌는데, 이 중 국내 학술지와 국외학술지에 각각 49건, 81건이 게재되었고, 연구기관 보고서는 91건이 게재되었다. 이러한 수치는 2007–2008년에 이루어진 체계적 고찰의 수보다 약 4배가 증가한 것으로, 체계적 고찰은 근래에 급격히 증가하였고, 최근에는 더 많아질 것으로 여겨진다. 하지만 의료 현장에서 보다 활성화되기 위하여서는 개선의 여지가 많으며 다음과 같은 과제가 놓여있다.

첫째, 우리나라에 필요한 건강문제에 대한 체계적 고찰이 증가하여야 한다. 우리나라에서 생성된 중요한 의료문제에 대

한 체계적 고찰은 아직 그 수가 적으며 따라서 활용할 수 있는 근거의 수준도 미약하다.

예를 들어, 서구에 흔한 유방암이나 대장암등에 대하여서는 다수의 코크란 리뷰가 있으나 우리나라를 비롯한 동양권에 많은 위암이나 자궁경부암에 관한 리뷰는 20%에 불과하다. 즉 대부분의 체계적 고찰이 외국의 건강문제를 다룬 것이기에 우리나라의 실정에 필요한 질문에 대한 해답을 주기 어려울 수 있으며 이를 극복하기 위하여서는 국내에 실제 적용할 수 있는 체계적 고찰을 시행하고 활용할 필요가 있다.

둘째, 체계적 고찰 결과가 현장에서 적극적으로 활용되어야 한다. 국내의 제반 의료 영역에서 체계적 고찰 연구가 널리 활용되고 있다고 할 수 없으며 의료인들 간에도 체계적 고찰에 대한 인식은 아직은 낮은 수준이며 임상현장에서 생성된 고찰을 사용하는 사람도 많은 편이라고는 할 수 없다. 따라서 학부나 전공의 등의 교육 기회를 늘리고 국내외 고찰 결과를 보다 가독성이 높게 전환하는 등 체계적 고찰의 생성과 활용에 힘을 필요가 있다.

셋째, 데이터베이스 등 관련인프라를 확충하여야 한다. 체계적 고찰을 돕고 그 결과를 활용하기 위하여서는 국내에서 생산된 체계적 고찰을 수집 정리하고, 체계적 고찰을 위한 임상시험 등 국내의 관련 자료를 데이터베이스로 구축할 필요성이 있다. 이러한 종류의 인프라를 구축하여야 한다.

넷째, 체계적 고찰 방법론 등에 대한 교육 기회 확대 등이 필요하다. 의료인을 대상으로 체계적 고찰 방법론을 교육하고 확산할 수 있는 기회의 확대가 필요하다. 그간 체계적 고찰 워크샵 등이 개최되었고 고찰방법에 대한 몇몇 지침서가 국내에서도 출간된 것은 고무적인 현상이지만 배전의 노력이 필요하다. 이에 관심을 갖는 연구인력도 아직은 소수이며 또 이 연구기회의 확대와 국내외 연구자들 간의 보다 토론의 장이 필요하다.

결론

체계적 고찰은 의학분야의 중요한 연구영역으로 자리잡았다. 최근 20여년간 논문의 양적인 면에서나 방법론적인

측면에서 많은 발전이 있어왔으며, 의학의 의사결정에서 중요한 역할을 하고 있다. 아울러 의학정보체계가 발전되면서 체계적 고찰을 수행할 수 있는 여건도 개선되었다. 한편 체계적 고찰의 타당성을 저하할 수 있는 연구의 이질성이나 출판비돌림의 문제는 앞으로 극복해야 하는 커다란 과제로 등장하고 있다.

우리나라에서도 활동이 늘고는 있으나 의료현장에서 활성화되기에는 아직 많은 과제가 놓여있고 의료인들은 현장에서 여러 제약요건으로 그 실현에 한계를 느끼곤 한다. 메타 분석이나 체계적 고찰을 시행하고자 하는 연구자에게는 보다 원활한 활동을 할 수 있는 여건이 마련되고 의료 현장에서는 그 산물을 충분히 이해하고 적용할 수 있는 여건이 필요하다.

찾아보기말: 체계적 고찰, 메타분석, 근거중심의학, 의학 문헌 검색, 출판비돌림

ORCID: Hyeong Sik Ahn: <http://orcid.org/0000-0002-2084-7466>

REFERENCES

- Report on certain enteric fever inoculation statistics. *Br Med J* 1904;2:1243-1246.
- Glass GV. Primary, secondary and meta-analysis of research. *Educ Res* 1976;5:3-8.
- Mulrow CD. The medical review article: state of the science. *Ann Intern Med* 1987;106:485-488.
- Cochrane AL. Effectiveness and efficiency: random reflections on health services. London: Royal Society of Medicine Press; 1999.
- Cochrane AL. 1931-1971: a critical review, with particular reference to the medical profession. In: Teeling-Smith G, Wells NE; Office of Health Economics. *Medicines for the year 2000*. London: Office of Health Economics; 1979. p. 1-11.
- Bero L, Rennie D. The Cochrane Collaboration. Preparing, maintaining, and disseminating systematic reviews of the effects of health care. *JAMA* 1995;274:1935-1938.
- Heejeng Son. Assessing methodology quality of Korean systematic review using AMSTAR. Seoul: Korea university; 2012.
- McAlister FA, Clark HD, van Walraven C, Straus SE, Lawson FM, Moher D, Mulrow CD. The medical review article revisited: has the science improved? *Ann Intern Med* 1999;131:947-951.
- Bravata DM, Olkin I. Simple pooling versus combining in meta-analysis. *Eval Health Prof* 2001;24:218-230.
- Grant MJ, Booth A. A typology of reviews: an analysis of 14 review types and associated methodologies. *Health Info Libr J* 2009;26:91-108.
- Higgins JP, Altman DG, Gotzsche PC, Juni P, Moher D, Oxman AD, Savovic J, Schulz KF, Weeks L, Sterne JA; Cochrane Bias Methods Group; Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011;343:d5928.
- Petitti DB. Meta-analysis, decision analysis, and cost-effectiveness analysis: methods for quantitative synthesis in medicine. New York: Oxford University Press; 1999.
- Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song F. Methods for meta-analysis in medical research. 1st ed. Chichester: Wiley; 2000.
- Egger M, Smith GD, Altman DG. Systematic reviews in health care: meta-analysis in context. London: BMJ Books; 2001.
- Glasziou P, Irwig L, Bain C, Colditz G. Systematic reviews in health care: a practical guide. Cambridge: Cambridge University Press; 2001.
- Stangl DK, Berry DA. Meta-analysis in medicine and health policy. Basel: Marcel Dekker; 2000.
- Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003;327:557-560.
- Colditz GA, Brewer TF, Berkey CS, Wilson ME, Burdick E, Fineberg HV, Mosteller F. Efficacy of BCG vaccine in the prevention of tuberculosis. Meta-analysis of the published literature. *JAMA* 1994;271:698-702.
- Dickersin K. The existence of publication bias and risk factors for its occurrence. *JAMA* 1990;263:1385-1389.
- Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R. Selective publication of antidepressant trials and its influence on apparent efficacy. *N Engl J Med* 2008;358:252-260.
- Yusuf S, Teo K, Woods K. Intravenous magnesium in acute myocardial infarction. An effective, safe, simple, and inexpensive intervention. *Circulation* 1993;87:2043-2046.
- ISIS-4: a randomised factorial trial assessing early oral captopril, oral mononitrate, and intravenous magnesium sulphate in 58,050 patients with suspected acute myocardial infarction. ISIS-4 (Fourth International Study of Infarct Survival) Collaborative Group. *Lancet* 1995;345:669-685.
- Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med* 2009;151:264-269.

24. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP, Clarke M, Devereaux PJ, Kleijnen J, Moher D. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Med 2009; 6:e1000100.
25. Sterne JA, Sutton AJ, Ioannidis JP, Terrin N, Jones DR, Lau J, Carpenter J, Rucker G, Harbord RM, Schmid CH, Tetzlaff J, Deeks JJ, Peters J, Macaskill P, Schwarzer G, Duval S, Altman DG, Moher D, Higgins JP. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. BMJ 2011;343:d4002.
26. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, Porter AC, Tugwell P, Moher D, Bouter LM. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Methodol 2007;7:10.

Peer Reviewers' Commentary

본 논문은 근거중심의료를 구현하는데 필수적인 요소인 체계적 문헌고찰의 개념과 실제 수행방법을 소개하는 내용으로 구성되어 있다. 체계적 문헌고찰은 급격히 증가하고 있는 의학논문들 가운데 연구주제에 맞는 논문을 체계적으로 검색하여 질 평가를 수행한 후 메타분석을 통하여 종합적인 결론을 이끌어냄으로써 과학적 근거를 중심으로 환자들에게 도움이 되는 의사결정을 내리는데 도움을 주고자 하는 연구방법이다. 체계적 문헌고찰에 관한 개념을 의사들에게 널리 알림으로써 근거중심의료가 국내에도 조속히 정착되고 활성화되는 계기를 만들 수 있을 것으로 기대되어 국내 의료수준 향상을 위해 도움이 되는 시의적절한 논문으로 판단된다.

[정리: 편집위원회]