

An Introduction of the Systematic Review and Meta-Analysis

Woo Jong Shin

Department of Anesthesiology and Pain Medicine, Hanyang University College of Medicine, Seoul, Korea

Systematic reviews and meta-analysis represent a specific type of medical research in which an original article is a unit of analysis. These methods of research are essential tools in integrating scientific information, increasing the internal and external validity of the conclusions of original articles, and suggesting areas for future research. Meta-analysis is becoming popular because it can combine results from similar studies to calculate an overall estimate of a treatment effect. They are also necessary for the practice of evidence-based medicine and the medical decision making. However, conducting good quality systematic reviews is not easy and difficult to interpret. Since analysis of a data with meta-analysis is a relatively new field of research, many clinicians have not had the opportunity to learn about systematic review and meta-analysis systematically. An introduction of the rationale for carrying out meta-analysis will be helpful to the clinician in interpreting the results of meta-analysis.

Key Words: Evidence-Based Medicine; Meta-Analysis; Review

Correspondence to: Woo Jong Shin
우471-701, 경기도 구리시 경춘로 153,
한양대학교 구리병원 마취통증의학과
Departments of Anesthesia and Pain Med-
icine, Hanyang University Guri Hospital,
153 Kyungchoon-ro, Guri 471-701, Korea
Tel: +82-31-560-2390
Fax: +82-31-563-1731
Email: swj0208@hanyang.ac.kr

Received 26 November 2014
Revised 24 December 2014
Accepted 31 December 2014

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.

서 론

1. 메타분석이란?

매년 많은 의학논문이 출판되고 있으므로 임상 의사들은 많은 연구정보를 요약하여 습득할 수 있는 능력이 필요하게 되었고, 일반 종설(narrative review)에서 생길 수 있는 비뚤림(bias)을 막기 위하여 여러 연구 결과를 요약하는 체계적인 방법의 개발이 필요하다. 근거중심의학(evidence-based medicine), 체계적 고찰(systematic reviews), 그리고 메타분석(meta-analyses) 등에 의해 의학의 진보와 혁신이 이루어지게 되었다. 체계적 분석은 “특정한 주제에 관하여 연관된 모든 연구물의 체계적 결합, 예리한 평가, 그리고 결합을 통하여 비뚤림을 줄이는 과학적인 전략의 적용(the application of scientific strategies that reduce bias by the systematic assembly, critical appraisal, and synthesis of all relevant studies on a specific

topic)”으로 정의될 수 있는데 재현 가능성이 있으며 명백한 방법론에 의해 확고한 연구방법과 목적으로 이루어진 문헌들의 종합이다. 이는 비뚤림과 확률적 오차(random error)를 줄이는 방법을 사용하여 다양한 연구의 결과를 통합한다. 모든 적절한 문헌을 검색하고, 연구물의 연구설계와 특성을 조사하고 자료를 종합하여 그 결과를 추정하는 것이다. 메타분석은 서로 다른 연구들의 양적인 효과 크기 결과들을 통계적으로 통합하기 위한 방법이다. 체계적 분석으로 각 연구 결과로부터 얻은 효과 크기(effect size)의 추정 값 즉 평균 차, 상대위험도(relative risk), 승산비(odds ratio) 등의 자료를 통계적 기법을 사용하여 결합하여 각 중재법의 가중치의 평균 요약 추정치를 얻은 뒤, 이 결과들을 분석하여 종합적인 결론을 얻는다. 그러므로 체계적 분석이 항상 메타분석을 포함하는 것은 아니다. 체계적 고찰을 하는 중에 메타분석 방법을 사용할 수도 있지만 통합하고자 하는 변수에 대한 충분한 정보가 각 연구문헌들에

서 제공되지 않거나 각 연구들의 동질성이 없는 경우에는 메타분석을 못할 수도 있다. 메타분석은 모든 과정을 재현할 수 있으며 대상이 되는 모든 연구들을 합하여 정밀한 결과를 추정할 수 있다는 장점이 있다. 그러나 연구 대상이나 치료중재법 그리고 결과변수에 대하여 대상 연구들 간에 충분한 동질성이 있는지를 사전에 검사한 뒤 통계적인 메타분석 사용여부를 결정하여야 한다. 비록 메타분석은 시간이 많이 소모되고 노동집약적이며 임상이나 통계의 전문가가 필요하다는 단점이 있지만 잘 계획되고 시행된 분석은 고급의 근거를 제공하는 유용한 도구가 될 수 있다.

2. 왜 메타분석인가?

무작위 대조군 연구(randomized controlled trials, RCTs)의 체계적 고찰과 메타분석은 근거중심의 기준(evidence-based guidelines)의 등급체계에 있어서 매우 높은 수준의 근거로 여겨지고 있다[1]. 메타분석은 비용이 많이 들어가는 대규모의 무작위 대조군 연구를 하기 곤란할 때 유용하다. 또한 어떤 경우에는 무작위 대조군 연구를 시행하는 것이 비윤리적이어서 많은 임상사들이 참여하기를 거부할 때 유용한 분석 방법이 될 수도 있다. 이런 경우에는 비교관찰연구(observational study)가 대신 사용할 수 있는 연구 방법이 될 수 있고, 이러한 경우에도 비교관찰연구로부터 얻은 자료를 이용하여 메타분석을 할 수도 있다. 그러나 이 경우에는 개별 일차 연구들에 내재된 비뚤림으로 인하여 메타분석이 매우 복잡하여 분석이 어려울 수도 있다. 비교관찰연구에서는 많은 교란변수로 인하여 효과크기가 왜곡될 수도 있기 때문이다. 비교관찰연구의 메타분석의 제약에 대해 많은 연구가 이루어졌다[2-4]. 어떤 연구자들은 메타분석이 연구물의 불완전한 문헌검색, 일차연구물의 선택과 포함기준의 비뚤림, 각 연구물의 이질적인 자료의 잘못된 결합, 비교관찰연구에서 시행한 메타분석, 분석결과 얻은 결과의 임상적용의 어려움 그리고 대상 연구물에 대한 지식의 부족 등을 이유로 메타분석을 비난하기도 하지만 이러한 단점에도 불구하고 수많은 메타분석 연구결과와 보고가 시간이 갈수록 증가하고 있다. 이 연구방법에 대한 관심이 증가함에 따라 여기서는 체계적 고찰과 메타분석에 대한 소개와 더불어 임상사들이 이러한 연구를 할 때 각 단계에서 생기는 문제들과 요점에 대해 살펴 보기로 한다.

시작하기 전에

먼저 체계적 고찰과 메타분석을 시작하기 전에 연구자는 반드시 기존에 같은 주제의 체계적 고찰이 수행되었는지 아니면 현재 진행 중인지 조사하여 알아보거나 하는 연구주제가 적당한지 알아본다. 만약 이미 같은 주제의 체계적 고찰이 있다면 새로 진행하려는 체계적 고찰의 방향을 위하여 충분한 질 평가가 이루어져야 한다. 일차 연구물을 선택 또는 배제하기 위한 투명하고 적절한 기준이

있어야 하며 각 연구물의 자료를 평가, 추출하여 통합하는 과정에서 비뚤림이 없어야 하고 투명하고 재현 가능성이 있어야 한다. 잘 구성된 평가계획서는 잘 맞추어진 연구 관심사, 문헌검색 전략, 일차연구물에서 요약하고 종합하는 자료의 종류, 그리고 양적인 통합의 요약된 자료 등을 포함하여야 한다. 이때 각 단계에서 생길 수 있는 여러 가지 예상하지 못한 비뚤림을 막을 수 있어야 한다. 연구방법을 사전에 미리 정하는 것이 고찰을 작성할 때 비뚤림이 생기는 것을 줄일 수 있다. 체계적 고찰의 결과를 보면 어떠한 치료의 중재법이 더 효과적인지 구별할 수 있고 더 신뢰할 수 있는 치료법으로 대체할 수 있을 것인지 알 수 있다[4]. 체계적 고찰의 목적은 첫째 방대한 양의 정보를 다룰 수 있는 크기 정도로 줄여서 일차 연구물들의 결과가 서로 일치하는지 아니면 결과를 일반화할 수 있는지 결정할 수 있도록 도와주고, 둘째로 연구결과의 출판과 새로운 효과적인 치료법의 시행 사이의 시간을 줄여준다. 셋째로는 각 연구물의 정보를 통합하여 전체 표본의 크기를 증가시켜 비뚤림을 감소시키고 연구의 신뢰도와 정밀도를 높이며 연구결과의 검증력을 배가시킨다. 새로운 연구를 시행하는 것 보다 체계적 고찰을 하는 것이 비용이 적게 들며 더 빠를 수 있다. Berman과 Parker 등은 “메타분석이 빠르지도 않고 쉽지도 않다(meta-analysis is neither quick nor easy)”라고 하여 메타분석은 간단하게 연구자료를 결합하는 방법이 아니라고 하였다[1]. 사전에 상당한 준비작업과 잘 정비된 연구계획이 개발되어야 메타분석이 의미가 있다. 그래서 각 연구 디자인에 따라 무작위 대조군 연구의 경우에는 “Preferred Reporting Items for Systematic review and Meta-Analysis” (PRISMA) [2], 그리고 비교관찰연구의 경우에는 “Meta-analysis of Observational Studies in Epidemiology” (MOOSE) [3]와 같은 준비계획서가 만들어졌으며 이 두 가지 기준을 참고로 하여 메타분석을 수행하는 데 필요한 각 단계를 살펴보고자 한다. 메타분석을 수행하기 위해서는 먼저 1) 구체적 질문의 설정(define the research question), 2) 문헌 검색(perform the literature search), 3) 일차 연구의 선택(select the studies), 4) 결과 자료의 추출(extract the data), 5) 자료의 분석(analyze the data), 그리고 6) 결과의 보고(report the results)와 같은 과정을 거치게 된다.

구체적 질문 만들기

구체적 질문을 만드는 목적은 기존에 존재하는 근거에 의해 결정해야 할 연구주제에 대한 결론을 얻기 위하여 작성한다. 질문은 네 가지 특성을 포함하고 있으며, 영문 머리글자를 따서 PICO라고 명명하였다. 환자 대상 문제(population; P), 치료 또는 중재(intervention; I), 비교자(comparator; C), 그리고 결과(outcomes; O)와 같은 어휘로 요약할 수 있으며 이러한 연구의문의 구성요소는 연구 디자인(study design; S)과 함께 일차 연구물을 선택할 때 사용될 수 있다

록 더 정교하게 만들어져야 한다. PICO 설정은 우선 ‘평가가 필요한 가’에 대한 답변을 얻기 위해 관련문헌을 찾는 과정에서 검색초안을 잡는다. PICOS의 설정은 연구의 범위를 정하는 데 중요한 영향을 미치므로 세심하고 정교하게 구성되어야 한다[5]. 계획서에 실제로 평가가 이루어지기 전에 구체적 질문이나 문헌선택 기준 등을 명확히 기술하고 수행과정을 자세히 제공하여야 비뮴림이 적은 근거를 얻을 수 있다. 평가계획서는 평가방법에 대한 구체적 내용이 기술되며, 주관성을 배제하기 위하여 평가가 수행되기 전에 작성되어야 한다. 계획서는 검증을 받아야 하며, 사전에 계획되지 못한 부분이나 과정 중에 수정 또는 추가 사항이 있는 경우에는 전문가의 의견을 반영하여야 하며 연구배경, PICOS, 평가방법, 검색어, 검색 데이터베이스, 출판형태, 연구유형, 언어, 문헌선택 및 배제기준, 질 평가 도구와 방법, 자료추출 및 전략 등과 같은 내용이 포함되어야 한다.

1. 연구 대상집단(Population)

포함된 연구 대상집단은 명확히 정의되어야 하고 연구대상의 건강 및 질환이 있는지 고려하여야 하며 대상 집단을 제한할 때는 명확하고 합리적인 기준을 제시하여야 한다. 초록에서 대상으로 하고 있는 군을 찾아 어떤 환자군까지 대상으로 할 것인지를 결정한다. 범위는 이질적인 연구를 통합하는 결과를 초래하지 않도록 한다.

2. 중재법과 대조군(Interventions and comparators)

시행되는 중재법에 대한 정확한 정의가 필요하며 최대한 상세히 기술한다. 중재법의 시술방법을 어떤 범위까지 포함해야 하는지 결정해야 한다. 중재법을 제한할 때는 명확하고 합리적인 기준을 제시한다. 대조군의 기술은 사전에 검토한 초록을 토대로 비교된 시술을 정리하고 현실을 고려하여 어떤 범위까지를 비교대상으로 해야 할지를 정해야 하며 구성요소를 최대한 자세히 기술하고 정의한다. 이에 따라 선택 배제의 범위가 달라진다. 처치를 하지 않은 경우에는 그 의미를 명확히 해주어야 한다.

3. 결과 평가변수(Outcomes)

핵심기술 자체의 결과와 해당 의료기술의 사용으로 인해 환자에게 미치는 결과를 구분하여야 한다. 해당 의료기술에 따라 측정되는 의료결과는 다르므로 전문가의 의견을 고려한다. 이에 따라 안전성과 유효성의 결과를 산출한다. 핵심적 결과 및 연관된 기존 문헌고찰에서 사용된 결과 항목의 사용을 고려한다. 사망률, 합병증을 등 안전성에 대한 자료와 생존율, 증상개선, 삶의 질, 만족도 등의 유효성 자료 등을 해당 의료기술별로, 초록을 토대로 정해야 한다.

4. 연구 디자인(Study design)

포함된 연구 디자인의 종류가 연구 결과의 신빙성과 효과 크기의 타당도를 결정하는 중요한 역할을 한다. 그러므로 연구 질문의 해

결에 가장 적합한 연구 디자인을 선택하고 디자인 선택의 논리를 제시한다. 이상적으로 치료적 시술에 대한 의료기술 평가는 무작위 대조군 연구를 포함하는 것이 가장 타당한 결과를 산출할 수 있는 연구유형이나, 의료기술 평가는 새로운 기술이 많으므로 해당 연구주제에서 무작위 대조군 연구가 부족하다. 그러므로 어떤 연구유형까지 평가에 포함시켜야 할 것인지 전문가와 함께 결정해야 한다. 특히 안전성은 증례연구를 통해 유용한 정보를 얻을 수 있다.

문헌검색

연구질문에 대한 PICO가 결정되면 구체적 질문에 대한 답을 주는 문헌을 찾고, 비뮴림을 제거하기 위해 계획단계에서 선택 및 배제기준을 각각 제시한다. 선택과 배제 기준은 논리적으로 서술되어야 하며 배제기준 중 하나라도 해당되면 그 문헌은 배제한다. 두 명의 평가자가 독립적으로 수행한 뒤 일치 여부를 확인한다. 사용할 데이터베이스로는 Scottish Intercollegiate Guidelines Network (SIGN) 기준에 따라, 체계적 문헌고찰을 위한 검색의 비뮴림을 최소화하기 위해 MEDLINE, EMBASE, CINAHL 및 Cochrane Library 데이터베이스를 검색한다. 이 단계가 메타분석에서 중요한 단계이다. 가능한 많은 연구물이 포함될 수 있도록 광범위한 검색을 한다. 연구의 포함기준은 연구 주제와 일치하는 모든 연구를 추출하는 것이다. 만약 연구의 포함 범위를 너무 광범위하게 설정하면 포괄적이며 외적 타당도가 증가하나 추출된 연구물이 방대하여 자료의 비교 합성에 어려움이 있으며 범위를 너무 축소하면 작성이 쉽고 읽기는 쉬우나 여러 개의 세부적인 검토가 필요하고 연구결과 일반화에 어려움이 있을 수 있다[6,7]. 검색전략은 가능한 한 투명하고 다시 재현할 수 있도록 서술되어야 한다. 검색어는 관련문헌에서 표현하는 용어를 통해 선정한다. 일반적으로 Patient 용어와 Intervention 용어를 함께 연구유형과 조합하여 검색전략을 수립한다. 일부 Comparator 용어도 같이 쓰기도 하나 일반적으로는 검색어로 사용하지는 않는다. 일반적인 데이터베이스로 검색하였을 때 정보가 충분하지 않은 경우 더 많은 정보를 수집하기 위해 수기 검색으로 검색하기도 한다. 수기검색은 찾은 전자문헌에서 인용한 참고문헌을 통하여 얻을 수 있으며 과학문헌 인용색인을 이용하여 검색할 수도 있다. 서지정보를 통해서도 출판정보가 나타나지 않아 확인이 어려운 문헌을 통칭하는 회색문헌(gray literature)으로는 학위논문, 진행중인 연구, 학회자료집, 그리고 연구보고서 등이 있다. 출판 비뮴림을 줄이기 위해 회색문헌의 검토여부를 결정한다. 또한 현재 진행되고 있는 연구를 확인한다.

비뮴림

이상적으로는 체계적 고찰은 출판 비뮴림을 피하기 위해 모든 연

구물을 포함해야 한다. 무작위 대조군 연구나 비교관찰연구 중 일부는 출판되지 못하여 문헌검색에 누락될 때 출판 비플림이 생길 수 있다. 출판되지 못한, 유의성이 부족했던 결과를 보인 연구물을 제외하고 분석하게 되는 경우, 훨씬 더 과장된 결과를 나타낼 수 있으므로 주의해야 한다[8]. 통계적으로 유의성이 있는 연구결과는 유의미한 연구결과로 인식되어 출판의도를 높이고 학술지의 게재 가능성이 높다[9-11]. 통계적으로 유의한 결과를 보인 자료는 출간되는 시간이 짧으며 검색될 확률도 높다[10]. 반면 출간되지 않은 연구물은 찾기가 어렵다. Easterbrook 등은 유의한 결과를 보인 비교관찰연구가 무작위 대조군 연구보다 더 많이 출간된다고 하였다[9]. 이러한 특성은 출간된 연구결과를 결합하는 메타분석에서는 분석 결과가 실제보다 과장되어 추정될 수 있다. 통계적으로 유의한 결과는 주로 영어로 출판되는 저널에 게재될 가능성이 높으며(언어 비플림: language bias), 다른 저널에 중복 출판될 가능성도 높다(중복출판 비플림: multiple publication bias). 또한 다른 논문에서 인용될 확률도 높다(문헌참고 비플림: citation bias)[11]. 이러한 비플림은 모두 출판 비플림과 관련된 특성으로 나타나며, 고찰의 결과에 영향을 미칠 수 있으므로 고찰의 결론과 추론을 이끌어 낼 때 반드시 고려해야 한다[8]. 영어를 공용어로 사용하지 않는 국가에서 수행된 연구물이 통계적으로 유의한 결과를 보인 경우 영어로 쓰인 저널에 더 잘 게재되는 경우에 언어 비플림이 생길 수 있다[12]. 그러므로 메타분석에서 영어로 쓰인 연구물만 포함하는 경우 결합효과의 크기가 왜곡되게 과장될 수 있다[12]. 메타분석에서 문헌검색을 영어로 출간된 연구로 제한하는 경우 또 다른 비플림이 생길 가능성이 있다. McAuley 등은 회색문헌을 제외하고 메타분석을 다시 할 때 평균 12%의 통합 효과크기가 영향을 받는다고 하였다[13]. Sutton 등은 Cochrane database에서 체계적 고찰을 분석한 결과 출판 비플림이 어느 정도 있다고 하였다[14]. 만약 저널들이 긍정적인 결과를 보인 연구물을 더 많이 게재하면 연구자들은 부정적인 결과를 보인 연구물들을 투고하지 않으려 할 것이다[9]. 또한 출간되지 못한 연구들은 연구 디자인이 부실하여 분석에서 제외되었을 가능성도 있을 수 있다. Easterbrook 등은 긍정적인 결과를 보인 연구가 그렇지 않은 연구보다 더 좋은 디자인을 보인다는 근거는 없다고 하였다[9]. 회색문헌과 출간되지 않은 연구를 찾기가 쉽지 않다. 연구기관의 사서정보 전문가의 도움으로 이러한 문헌을 찾을 수 있다. 깔때기 그림(funnel plot)은 가로축에 일차 연구로부터 얻은 중재효과의 추정치를 설정하고 세로축에는 추정치의 정밀도를 나타내는 표준오차의 역수 또는 연구 표본수 등을 설정하여 시각적으로 그려진 산점도(scatter plot)를 의미하는데 이를 통해 시각적으로 판단하는 방법이 출판 비플림을 평가하는 데 도움이 된다. 깔때기 그림이 대칭일 경우 출판 비플림의 가능성은 줄어들며, 비대칭일 경우 출판 비플림의 가능성이 높아진다. 그러나 비대칭적이라고 하여 반드시 출판 비플림이 있음을 의미하는 것이 아니라 선택 비플림, 연구의 질이 낮

은 경우, 이질성, 부적절한 분석 등과 같은 다른 원인으로 인해 발생할 수도 있다. 표본수가 적은 연구들은 그래프의 아래에 넓게 분포되고, 표본수가 많은 연구들은 깔때기 위쪽의 좁은 부분에 분포된다. 그러므로 깔때기 그림은 다양한 표본수를 가진 다수의 연구들이 필요하다는 한계가 있다. 현재 진행되고 있는 연구를 찾는 것은 후속연구와 연구의 보완을 위해 유용한 근거를 제공한다는 점에서 매우 중요하다. 연구자들은 “www.ClinicalTrials.gov”이나 “www.who.int/trialsearch”와 같은 웹사이트를 방문하여 현재 진행되고 있는 연구를 찾을 수 있다. 또한 일반 인터넷에서도 출간되지 않은 보고서, 또는 학회초록과 같은 회색문헌을 찾을 수 있다. 방대한 양의 출판된 서적을 조사하는 것도(text mining) 유용한 접근법이다. 현재 탐색 중이며 조만간 관련 연구를 찾는 데 중요한 부가적 방법이 될 것이다. 출판된 서적을 찾는 것은 외견상 관련이 없는 사실이나 새로운 아이디어나 가설을 만드는 연결고리가 될 것이며 이런 과정을 통해 비플림이나 오류를 감소시킬 것이다.

문헌의 선택과 질 평가

문헌 검색이 완료되면 각 연구들은 포함기준에 따라 분류한다. 포함 배제 기준은 메타분석 초기에 정의되어야 하며 연구의 선택은 주제에 맞는 연구를 모두 포함시키는 것이다[15,16]. 구체적 질문에 대한 답을 주는 적절한 문헌을 찾기 위하여 명확한 선택과 배제 기준을 각각 제시하여야 하며 목적에 따라 논리적으로 구성하여야 한다. 문헌을 선택할 때 설정된 배제기준 중 하나라도 해당되면 해당 문헌은 배제해야 하며 선택기준은 모두 만족해야 한다. 연구의 선택은 두 단계를 거치는데 첫째로 제목과 초록을 보고 선택한다. 둘째로 제목과 초록만으로는 결정을 못하는 경우 본문을 보고 결정을 한다. 한 명의 평가자가 연구를 선택할 경우 약 8%의 누락이 있을 수 있는 데 비해 두 명의 평가자가 독립적으로 수행할 경우에는 해당하는 모든 연구물을 찾을 수 있다고 한다[17]. 두 명의 평가자가 참여하여 독립적으로 수행한 뒤 일치(inter-assessor reliability) 여부를 kappa 통계를 통하여 확인하고 전문가의 자문과 함께 최종적으로 문헌을 선택한다[18]. 비플림은 결과의 신뢰도에 영향을 줄 수 있으므로 연구의 질(quality)을 평가하여야 한다. 연구의 질이란 연구 설계, 수행, 분석, 임상적 타당성(clinical relevance) 및 보고의 질과 관련이 있는 다차원적 개념이다. 질 평가 과정은 각 논문이 주는 정보가 객관적이고 타당성 있는지를 알아보는 과정이다. 모든 논문의 주장이 다 객관적인 근거가 있는 것은 아니므로 객관적인 결과에 영향을 주는 비플림이 개입될 가능성이 있다. 질 평가는 주로 연구유형에 초점을 두며 연구결과의 타당성에 영향을 미치는 비플림이 얼마나 존재하는지에 의해 결정된다. 그러므로 비플림에 의해 중요한 중재효과가 모호해지기 쉽다. 포함된 연구의 장, 단점을 기록하여 연구의 결과가 연구 디자인이나 수행에 의하여 영

향을 받는지 알아볼 수 있다. 내적 타당성이란 임상연구에서 체계적 오류(systematic error)를 최소화한 정도를 말한다. 내적 타당성을 위협하는 비뚤림은 선택 비뚤림, 실행 비뚤림, 탈락 비뚤림, 결과 확인 비뚤림 등이 있다. 메타분석의 타당성은 포함된 연구의 질에 좌우되므로 질 평가는 반드시 필요한 과정이라 할 수 있다. 연구자들은 가능한 많은 연구물을 포함시키기를 원하지만 질이 낮은 연구는 메타분석에서 제외하는 것이 연구자로 하여금 적은 자료로 더 완벽한 분석을 할 수 있게 한다[19]. 질 평가는 근거의 강점을 제공할 뿐 아니라 후속 연구에 필요한 기준을 제공한다. 질 평가는 연구의 결과가 치료나 예방, 진단과 정책의 결정에 충분히 방향을 제시할 수 있는지 해답을 얻는 데 도움을 준다. 또한 질 평가는 연구 목적에 맞는 디자인의 적절성, 비뚤림의 위험도, 연구의 질에 관련된 여러 가지 이슈, 측정결과와 선택, 통계적인 문제, 중재와 보고의 질, 그리고 일반화 등에 관한 내용 등을 포함하여야 한다. 이러한 질과 관련된 여러 문제의 중요성은 고찰의 성격과 핵심내용에 따라 다르다. 무작위 대조군 연구의 질을 평가하는 여러 종류의 방법이 개발되어 왔다[20]. 점검목록 방식(checklist)은 무엇을 보고해야 하는지에 대한 지침서 역할을 하는 반면 척도방식(scale)은 연구에서 비뚤림을 정량화하는 데 도움이 된다. 즉 연구의 특성을 점수화하는 반면, 점검목록 방식은 점수와는 상관이 없다. 비록 질 평가가 필요하지만 이러한 점검목록 방식이나 척도방식을 사용하는 데 주의를 해야 한다[20,21]. Moher 등은 체계적 고찰에서 질은 연속적이지 않으므로 독자들이 무비판적으로 받아들이지 말 것을 제안하였다[22]. Delaney 등은 중환자 논문에서 메타분석의 질을 체계적으로 평가하여 전반적인 질이 낮다고 보고하였으며 임상사들은 이러한 연구결과들을 임상에 적용하기 전에 연구물들을 평가해 보아야 한다고 하였다[23]. 비록 체계적 고찰과 메타분석이 폭발적으로 증가하고 있지만 이러한 연구의 질과 관련한 연구들을 보면, 모든 체계적 고찰이 반드시 체계적이지 않다는 것을 알 수 있다[3,24]. 더욱이, 체계적 고찰의 질은 매우 다양하여 방법론적인 질 평가는 반드시 필요할 뿐 아니라 필수적으로 해야 한다[25,26]. 이러한 비판적이고 광범위한 고찰이 있을 뒤 새로운 지침(A measurement tool to assess the methodological quality of systematic reviews, AMSTAR) 체계적 고찰을 평가하는 가장 적합한 가이드라인으로 개발되었다[27]. 이런 질과 관련된 내용은 전문가나 고찰을 수행하는 인적자원에 의해 많이 달라진다. 대부분의 연구 목적은 의료기술의 효과와 합병증을 포함하여야 하는데 일반적으로 무작위 대조군 연구는 안전이나 합병증에 관한 내용을 자세히 보고하지는 않고 비교관찰연구에서 더 자세히 보고되고 있다.

자료의 추출

자료 추출은 선택된 논문에 대해 연구의 특징과 결과물에 대한

필요한 정보자료를 체계적으로 추출하고 정리하는 단계이다. 연구자는 의료기술의 목적과 평가의 형식에 맞는 자료 추출 양식을 개발하여 얻고자 하는 정보의 양이 너무 자세하지 않고, 또 너무 간결하지도 않게 중요한 자료를 추출하여야 한다. 동일한 표본으로서 다른 연구자에 의해 검토를 한 뒤 사용한다. 연구자는 필요한 정보가 무엇인지 고려하여 분석에 요구되는 자세한 정보 및 자료를 추출한다. 일차 연구물의 결과 크기와 형식이 서로 달라서 자료 추출에 있어 어려움이 있을 수 있으며, 자료를 합성할 때 약간의 변형이 필요할 수 있다[28]. 만약 자료의 합성이 어려운 경우에는 합성을 그만두고 체계적 고찰로 분석을 종료할 수 있다. 일반적으로 자료추출의 정확성과 일관성을 유지하는 것이 주관적인 판단으로 인한 실수를 줄일 수 있다. 두 명의 연구자에 의해 독립적으로 수행되어야 하며 내용이 일치하지 않을 경우 토론을 통해 상호 협의하고, 합의되지 않을 경우 전문가의 도움이 요구된다.

자료의 합성과 분석

선택된 논문에서 얻은 결과로 양적인 합성을 하고, 그 결과를 제시하는 단계로 통합하는 방법이나 연구 디자인 및 결과의 형태에 따라 다를 수 있다. 기술적 결과 합성은 각 문헌의 의료결과를 기술하고 해석의 유의점, 근거의 타당성, 이질성 등을 기술한다. 얻은 자료를 합성할 때 일차 연구들에서 얻은 치료중재의 효과가 연속성이 있는지, 불연속성의 가능한 원인이 무엇인지 등을 알아보아야 한다. 이 단계는 개별연구를 명확하게 서술적으로 요약하여야 할 수가 있다. 즉 서술적 고찰이 체계적 분석에서 반드시 필요한 이유이다. 메타분석에서 각 일차 연구결과들의 합성은 중재효과를 추정하는 정밀도와 검정력을 증가시킨다. 대부분의 메타분석은 먼저 각 연구의 결과를 먼저 계산하고 분석한 뒤 일차 연구물의 통계치를 결합하여 전반적인 효과의 결합값을 얻는다. 이때 효과크기의 정밀도와 역분산(inverse of the variance of the effect estimate)에 비례하는 가중치를 사용하기도 한다[1,19]. 그러므로 대규모의 연구는 소규모의 연구에 비해 결합값에 더 많은 영향을 미칠 수 있다[29]. 각 연구의 가중값에 의해 얻은 결합값은 더욱 믿을 만하다. 양적인 합성결과는 신뢰구간(confidence intervals)과 함께 점 추정치(point estimates)로 표시하거나 정확한 P값으로 표시한다. 민감도 분석(sensitivity analyses)은 메타분석 결과가 분석에 사용된 특정 방법들이나 각 연구에서 얻어진 결론들에 얼마나 영향을 받는지를 평가하기 위하여 주된 메타분석 결과에서 특정자료나 분석방법에 변화를 준 뒤 다시 분석하여 얻은 결과의 견고함(robustness)을 살펴봐야 결과에 영향을 미치는지 알아보는 방법으로 얻은 자료의 임상적 통계적인 이질성이 의심될 경우 할 수 있다[30].

**고정- 또는 변량효과모형
(Fixed-Effect and Random-Effect Model)**

메타분석을 수행할 때 두 가지 분석모형에 따라 분석방법을 다르게 할 수 있다. 고정효과모형(fixed-effect model) 방법은 모든 연구들은 동일한 중재효과(effect of treatment)를 가지고 있는데 연구결과들 간에 변량(variation)이 나타나는 이유는 단지 연구들의 표본추출변동(random error) 때문이라고 가정하는 모형이다. 따라서 각 연구들의 중재효과들 간에 보이는 변동(variability in results within-study)의 크기가 단순한 표본추출오차에 의한 것이라고 믿어지는 경우 즉 각 연구들의 연구 디자인이나 방법 등이 서로 비슷한 경우에 사용되는 모형이다. 이 방법으로 분석하는 경우에는, Mantel-Haenszel 방법, inverse variance 방법, 그리고 Peto 방법을 선택할 수 있으며 각 연구에 적합한 방법을 선택한다. Fig. 1은 이 모형의 예를 보여준다. 변량효과모형(random-effect model)은 결과를 합성하고자 하는 연구들 사이에 이질성(heterogeneity)이 있다는 가정하에 분석하는 모형이다. 각 효과크기들 사이에 관찰되는 변동은 표본추출변동(variability in results within-study)과 각 연구들 사이의 변동(variability in results between-study)이 같이 표현된 것이라고 가정한다. 연구들 사이에 설명할 수 없는 이질성이 있다는 근거가 있을 경우에 사용하며 비록 이질성 검사 결과 유의하지 않은 결과가 나왔더라도 연구들이 동질하지 않다고 가정할 때 이 방법을 사용한다. Fig. 2에서 두 가지의 변동을 볼 수 있다.

이질성 조사(Exploring Heterogeneity)

이질성이란 결합하고자 하는 일차 연구들의 다양성으로 인하여 생기는 차이를 의미한다. 임상적 또는 연구 방법론적인 다양성으로 인해 나타난다. 임상적인 원인은 연구가 이루어진 지역이나 연구 대상의 특성, 연구에 이용된 중재법의 양이나 강도, 또는 의료 결과의 차이 등을 의미하며, 방법론적인 다양성은 연구 디자인과 질 그

리고 결과의 분석 차이 등에 기인한다. 만약 동질하지 않다면 결합된 효과추정치는 분석하기 곤란하므로 이질성의 원인을 알아야 하며 원인을 밝힐 수 없을 경우 결합된 요약 추정치를 제시하지 말아야 한다. 이질성을 확인하는 방법으로는 첫 번째, 숲 그림(forest plot)을 통한 시각적 방법과 두 번째, 각 연구들의 중재효과가 공통 중재효과 값으로부터 얼마나 차이가 나는지를 검정하는 방법인 Cochran's Q-test (chi-square test) 방법이 있다. 이 방법은 대부분의 연구 대상이 변량이 존재하지 않으며 동질하다는 귀무가설(null hypothesis)에서 시작하며 이를 기각할 수 있는 통계적인 power가 낮다. 즉 비록 연구 대상들 사이에 이질성이 존재하더라도 이를 기각할 수 있는 일반적인 유의수준에서는 유의하지 않은 결과를 나타낼 수 있으므로 유의수준을 좀 더 크게($P < 0.1$) 정하여야 한다. P 값이 0.10 이상일 경우 고정효과모형으로 메타분석을 수행하며, 0.10 미만일 경우, 변량효과모형을 고려할 수 있다. 그리고 마지막으로 비일관성을 정량화시킨 Higgins I^2 -statistic 방법이 있다. I^2 값이 25%까지는 이질성이 낮다고 하며, 50%까지는 이질성이 중간 정도이고, 75%까지는 "이질성이 높다"라고 해석한다[31]. 그러나 이질성 정도의 해석은 효과크기의 방향과 정도 그리고 이질성에 대한 근거의 강도에 따라 달라진다. 만약 통계적인 이질성이 관찰되면, 변량의 가능한 원인을 반드시 찾아야 할 것이며 연구들을 결합할 것이 적절한지 또는 어떻게 할 것인지 결정해야 한다[32]. 가장 좋은 방법은 항상 변량효과모형을 사용하거나 아니면 두 가지 모형을 모두 사용한 뒤 치료효과의 결과를 비교하여 보는 것이다. 메타회귀분석(meta-regression) 방법이 치료효과의 추정에 필요한 연구들의 특성을 조사하는 데 사용할 수 있다[33]. 메타회귀분석은 연구수준의 공변량에 대한 연구 요약추정치의 회귀분석이다. 따라서 분석단위는 개별 연구이다. 치료효과와 조사된 연구특성 간에 선형관계가 있다면 메타회귀분석은 하위그룹 분석보다 통계적 검정력이 더 크다. 메타분석과 마찬가지로 대규모 연구는 소규모의 연구보다 연구효과의 정밀도에 더 큰 영향을 주어 가중되는 회귀값을 나타낸다. 하위그룹 분석은 연구들 간의 이질성이 임상적인 원

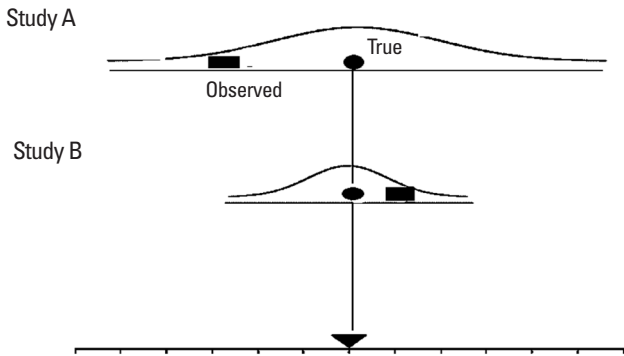


Fig. 1. Fixed-effect model; distribution of random error.

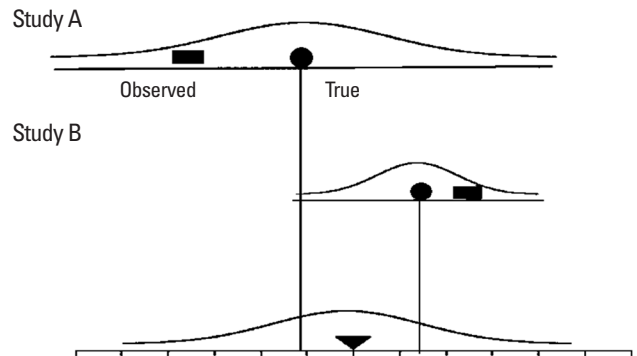


Fig. 2. Random-effect model; between study and within study variance.

인이나(성, 연령, 연구가 이루어진 지역이나 연구대상의 특성, 연구에 이용된 중재법의 양이나 강도, 질병의 중증도, 진단기준, 결과변수의 정의, 추적 기간 또는 의료 결과의 차이 등) 방법론적인 다양성(연구 디자인과 질, 무작위 방법의 차이, 연구 질의 차이, 그리고 결과의 분석 차이 등) 등에 의해 기인한 것인지를 파악하고자 환자군이나 연구방법의 특성이 비슷한 연구들을 소그룹으로 만들어 분석하는 방법을 말한다. 하위그룹 분석은 포함된 연구의 수가 작아지므로 검정력이 낮아져 위 양성이나 위 음성의 결과의 가능성이 증가하므로 해석에 신중하여야 하며, 분석결과는 연구주제에 대한 결론을 내리기 위한 것이 아니라 이질성의 원인을 찾고 후속 연구를 위한 새로운 가설을 제시하기 위한 용도로 사용한다.

결과의 보고

연구자는 연구의 결과를 투명하게 보고해야 한다. 그러나 연구방법과 대상, 연구의 수행 및 분석방법이 명료하게 보고되지 않은 경우가 많다. 정확하지 않은 보고는 결과의 해석을 어렵게 할 뿐 아니라 문헌평가에서 중요한 세부항목이 누락된 경우는 결점이 많은 방법론으로 인해 부적절한 것으로 간주된다. 여러 연구그룹에서 연구의 질을 향상시키고자 연구 디자인별로 꼭 포함되어야 할 항목들을 정리하여 표준화된 보고지침을 제공하였다. 메타분석의 수행과 무관하게 체계적 고찰을 보고하는 흐름도(flow chart)와 권고안에 예전에는 quality of reports of meta-analyses of randomized controlled trials (QUOROM) [5] 그룹이었던 PRISMA [2] 그룹에 의해 개발되었다. 가독성이(readability) 보고의 가장 중요한 점이며 독자들이 실제로 읽는 유일한 부분이 요약문이나 초록이므로 보고에서 가장 중요한 부분이 되었다. 무작위 대조군 연구의 메타분석에 대한 자세한 보고지침이 PRISMA 권고안에 기술되어 있다[34]. 제목에는 반드시 메타분석임을 표시하여야 하고 방법에서는 문헌검색과 사용한 데이터베이스도 언급하여야 한다. 문헌의 선택, 질 평가, 자료의 추출 방법과 합성도 기술한다. 결과항목에선 연구 선택의 각 단계에서 포함된 연구를 보여주는 흐름도를 제시하여야 하고

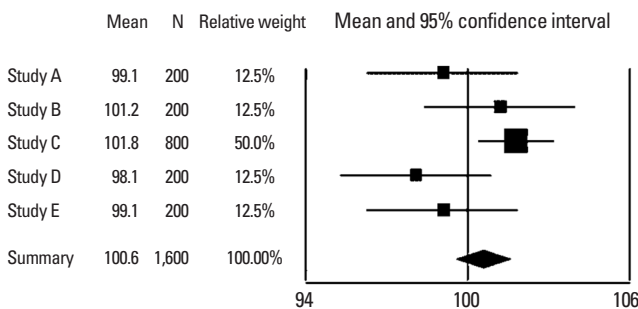


Fig. 3. Example of a fixed-effect analysis.

각 연구의 결과를 보여주는 숲 그림으로 이질성이 있는지 제시하고 결합된 요약 통계치를 95% 신뢰구간과 함께 표시한다. 만약 민감도 검사나 메타회귀분석을 하였으면 함께 표시한다. 토의 항목에선 주요결과와 가능한 비뿔림과 이질성의 원인을 기술한다. 숲 그림은 추정되는 효과크기와 신뢰구간, 그리고 요약 추정치를 함께 표시하는 중요한 그림이다. 각 연구는 크기와 특성에 따라 구분하여 있어서 독자들로 하여금 이질성을 알 수 있도록 한다. 만약 신뢰구간이 겹치지 않는다면 연구들 사이에 변량이 상당히 있어서 메타분석이 어려울 수도 있다. 이런 경우에는 변량의 원인을 조사하여 환자와 연구결과 사이에 관계를 추정하는 것이 필요하다. Fig. 3과 4에서 각각 고정- 또는 변량효과모형의 분석의 사례를 보여주며 각 연구의 효과 크기와 신뢰 구간을 보여준다. 각 연구의 효과크기를 x축에, 그리고 정밀도를 y축에 표시한 산점도를 보여주는 깔때기 그림은 출판 비뿔림을 평가하는 데 사용한다[19,35]. 전체적인 표본의 크기를 y축에 표시할 수도 있으나 표준오차의 역의 값도 많이 사용된다[14]. 만약 출판 비뿔림이 없다면 산점도 그림은 깔때기를 뒤집은 모양이 될 것이다. 표본수가 작은 연구는 그림에서 깔때기의 밑부분이 넓게 분포하고 표본의 크기가 크면 윗부분에 모이는 효과추정치의 모습을 보일 것이다. 만약 부정적인 결과를 보인 표본 크기가 작은 연구는 출판되지 않을 것이며 그림의 왼쪽 밑부분

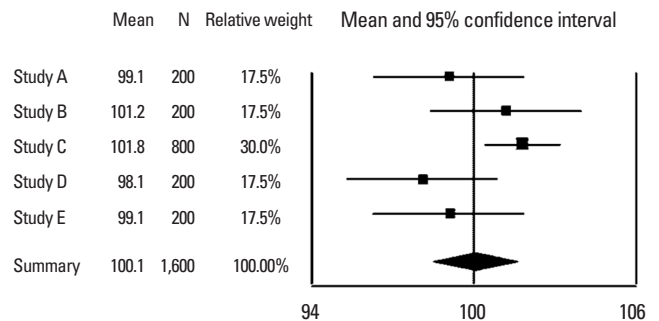


Fig. 4. Example of a random-effect analysis.

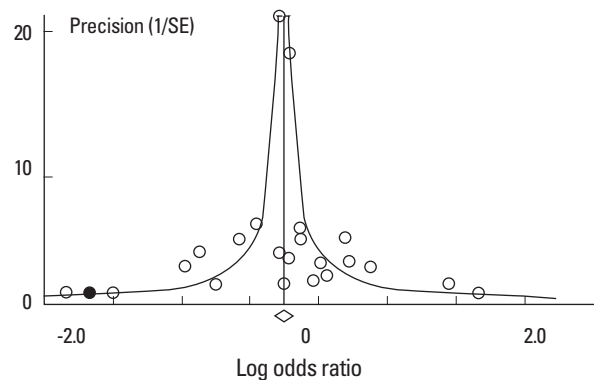


Fig. 5. Example of a funnel-plot of precision by log odds ratio.

에 자료의 점이 없는 비대칭적인 그림을 나타낼 것이다. Fig. 5는 이론적인 깔때기 그림의 예를 보여주고 있다. 비록 메타분석이 근거의 최상위급의 연구로 여겨지고 다른 연구설계보다 많이 인용되지만 여전히 잘 수행된 무작위 대조군 연구와 비교할 때 타당성에 관한 의문이 여전히 남아있다[36]. 그러나 메타분석의 결과는 치료효과를 시험하기 위하여 후속적인 대규모 무작위 대조군 연구를 계획하는 데 유용하나 대규모 후속연구와 결과를 비교하면 일치하는 정도가 비교적 낮은 편이다[37]. 그렇다고 메타분석의 결과를 신뢰하지 못하고 메타분석을 배척해야 한다는 것을 의미하지는 않는다. 오히려 메타분석과 관련하여 제한점과 비뚤림을 찾아내어 비뚤림이 적은 잘 계획된 메타분석을 할 필요성을 보여준다.

결론

지금까지의 내용은 메타분석의 연구에 관심이 있는 임상 의에게 실질적인 내용을 보여주었다. 아직까지 우리나라에서 메타분석은 일부에서 연구물을 내는 초보단계이며 각 전문분야에서 광범위하게 연구가 시행되지는 않고 있는 실정이다. 메타분석은 때로는 본래 가지고 있는 한계 때문에 비판을 받지만 그러한 비판은 메타분석 그 자체의 문제가 아니라 주로 방법의 실행 중에 생기는 문제이다. 메타분석은 비뚤림이 있는 결론을 피하기 위해 연구의 질 평가와 더불어 엄격한 방법을 적용하여 수행된다. 메타분석은 적절하게 적용이 되었을 경우 각 연구들로부터 얻은 자료를 확실하고 명료하게 결합시키는 장점이 있으며 또한 결합에 사용되는 강력한 도구로서 연구들에서 얻은 자료로부터 잘못된 결론을 내리는 것을 피하고 의미 있는 결론을 도출하는 데 많은 도움을 준다. 궁극적으로 특별한 주제에 대하여 환자에 관한 결정을 하는 것은 임상 의사이므로 메타분석은 근거의 양 및 질과 일관성, 연구결과의 일반화 가능성, 대상집단에서의 직접 적용 가능성, 임상적 영향, 수행가능성 등의 내적 타당도를 분석한 자료를 바탕으로 해당 중재기술의 안전성 및 유효성에 관한 최종 결정을 하는 데 보다 더 정밀한 도구로서 사용될 수 있다.

REFERENCES

- Berman NG, Parker RA. Meta-analysis: neither quick nor easy. *BMC Med Res Methodol* 2002;2:10.
- Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *J Clin Epidemiol* 2009;62:e1-34.
- Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *JAMA* 2000;283:2008-12.
- Petrie A, Bulman JS, Osborn JF. Further statistics in dentistry Part 8: systematic reviews and meta-analyses. *Br Dent J* 2003;194:73-8.
- Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of Reporting of Meta-analyses. *Lancet* 1999;354:1896-900.
- Horwitz RI. "Large-scale randomized evidence: large, simple trials and overviews of trials": discussion. A clinician's perspective on meta-analyses. *J Clin Epidemiol* 1995;48:41-4.
- Eysenck HJ. Meta-analysis and its problems. *BMJ* 1994;309:789-92.
- Song F, Eastwood AJ, Gilbody S, Duley L, Sutton AJ. Publication and related biases. *Health Technol Assess* 2000;4:1-115.
- Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. *Lancet* 1991;337:867-72.
- Stern JM, Simes RJ. Publication bias: evidence of delayed publication in a cohort study of clinical research projects. *BMJ* 1997;315:640-5.
- Egger M, Smith GD. Bias in location and selection of studies. *BMJ* 1998;316:61-6.
- Egger M, Zellweger-Zahner T, Schneider M, Junker C, Lengeler C, Antes G. Language bias in randomised controlled trials published in English and German. *Lancet* 1997;350:326-9.
- McAuley L, Pham B, Tugwell P, Moher D. Does the inclusion of grey literature influence estimates of intervention effectiveness reported in meta-analyses? *Lancet* 2000;356:1228-31.
- Sutton AJ, Duval SJ, Tweedie RL, Abrams KR, Jones DR. Empirical assessment of effect of publication bias on meta-analyses. *BMJ* 2000;320:1574-7.
- Slavin RE. Best evidence synthesis: an intelligent alternative to meta-analysis. *J Clin Epidemiol* 1995;48:9-18.
- Oxman AD, Guyatt GH. The science of reviewing research. *Ann N Y Acad Sci* 1993;703:125-33; discussion 33-4.
- Edwards P, Clarke M, DiGuseppi C, Pratap S, Roberts I, Wentz R. Identification of randomized controlled trials in systematic reviews: accuracy and reliability of screening records. *Stat Med* 2002;21:1635-40.
- Kraemer HC. Correlation coefficients in medical research: from product moment correlation to the odds ratio. *Stat Methods Med Res* 2006;15:525-45.
- Sutton AJ, Abrams KR, Jones DR. An illustrated guide to the methods of meta-analysis. *J Eval Clin Pract* 2001;7:135-48.
- 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.
- Juni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA* 1999;282:1054-60.
- Moher D, Tetzlaff J, Tricco AC, Sampson M, Altman DG. Epidemiology and reporting characteristics of systematic reviews. *PLoS Med* 2007;4:e78.
- Delaney A, Bagshaw SM, Ferland A, Manns B, Laupland KB, Doig CJ. A systematic evaluation of the quality of meta-analyses in the critical care literature. *Crit Care* 2005;9:R575-82.
- Jadad AR, Cook DJ, Jones A, Klassen TP, Tugwell P, Moher M, et al. Methodology and reports of systematic reviews and meta-analyses: a comparison of Cochrane reviews with articles published in paper-based journals. *JAMA* 1998;280:278-80.
- Mulrow CD. The medical review article: state of the science. *Ann Intern Med* 1987;106:485-8.
- McAlister FA, Clark HD, van Walraven C, Straus SE, Lawson FM, Moher D, et al. The medical review article revisited: has the science improved? *Ann Intern Med* 1999;131:947-51.
- Shea BJ, Hamel C, Wells GA, Bouter LM, Kristjansson E, Grimshaw J, et al. AMSTAR is a reliable and valid measurement tool to assess the meth-

- odological quality of systematic reviews. *J Clin Epidemiol* 2009;62:1013-20.
28. Egger M, Schneider M, Davey Smith G. Spurious precision? Meta-analysis of observational studies. *BMJ* 1998;316:140-4.
 29. Egger M, Smith GD, Phillips AN. Meta-analysis: principles and procedures. *BMJ* 1997;315:1533-7.
 30. Thompson SG. Controversies in meta-analysis: the case of the trials of serum cholesterol reduction. *Stat Methods Med Res* 1993;2:173-92.
 31. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002;21:1539-58.
 32. Thompson SG. Why sources of heterogeneity in meta-analysis should be investigated. *BMJ* 1994;309:1351-5.
 33. Greenland S, Salvan A. Bias in the one-step method for pooling study results. *Stat Med* 1990;9:247-52.
 34. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009;6:e1000097.
 35. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997;315:629-34.
 36. Davey Smith G, Egger M. Meta-analysis. Unresolved issues and future developments. *BMJ* 1998;316:221-5.
 37. LeLorier J, Gregoire G, Benhaddad A, Lapierre J, Derderian F. Discrepancies between meta-analyses and subsequent large randomized, controlled trials. *N Engl J Med* 1997;337:536-42.