

- 초청의 글 -

체계적 문헌고찰 연구의 비평적 평가: A MeaSurement Tool to Assess systematic Reviews 2를 중심으로

서현주

충남대학교 간호대학

Critical Appraisal for Systematic Review: Focused on A MeaSurement Tool to Assess Systematic Reviews 2

Seo, Hyun-Ju

College of Nursing, Chungnam National University

Researchers and healthcare professionals need to assess the confidence of review results before accepting results of a single systematic review, considering its reputation and influence based on the hierarchy of evidence in an evidence-based practice. A MeaSurement Tool to Assess systematic Reviews 2 (AMSTAR 2), an updated version of AMSTAR, was published in 2017, consists of 16 domains, and is useful as a critical appraisal tool for systematic review of randomized or non-randomized studies of healthcare interventions, or both, in terms of inter-rater agreement and intuitiveness of items.

Notably, AMSTAR 2 was not designed to generate an overall score. Moreover, the assessors should evaluate overall confidence in the results of systematic review, by determining critical domains and taking account of the potential impact of critical flaw of each domain on the confidence of the review results.

Key words: Critical appraisal, Systematic review, Methodological quality

들어가며

방대한 양의 일차연구를 포함한 연구문헌들이 출판되고 있기 때문에, 임상실무를 포함한 정책결정과정에서 보건의료 의사결정자들은 체계적 문헌고찰을 근거기반 보건의료를 달성하기 위한 핵심 도구 중의 하나로 사용해오고 있다.¹⁾ 연구질문에 관한 최상의 이용가능한 근거를 검색, 선정, 비평, 요약하는 체계적 문헌고찰은 근거중심 보건의료의 철학적 관점에서, 근거의 위계에서 최상위에 위치하여 황금기준으로 간주된다.²⁾ 체계적 문헌고찰의 결과들은 신뢰할만한 임상실무지침(임상실무지침내의 권고안) 개발 및 임상적 의사결정에 정보를 제공하기 위한 가장 타당한 근거를 제공한다.³⁾

그러나 최근 불필요하고 오해의 소지가 있으며 이해상충이 있는 체계적 문헌고찰 및 메타 분석이 대량 생산되고 있다는 보고가 있었다. 구체적으로, 약 10~20%의 체계적 문헌고찰만이 리뷰프로토콜을 PROSPERO와 같은 등록처에 등록하거나 저널에 출판하였으며, 무작위대조연구만을 포함시킨 메타분석연구의 가장 흔한 결론은 근거가 약하거나 불충분하여 고찰의 결과는 환자 치료나 보건정책적 측면에서 최상의 증거가 무엇인지에 대한 정보를 제공하고 있지 않았다. 방법론적으로 엄격하고 정보를 제공해주는 메타분석은 단지 3%에 지나지 않았다. 이러한 체계적 문헌고찰들은 종종 쉽게 생산되고 출판 가능한 성과물이나 제약회사나 제조회사의 마케팅 도구로 사용되기도 한다는 것이다.⁴⁾ 즉, 비체계적인 방법으로 수행된 체계적 문헌고찰 및 메타

주요어: 비평적 평가, 체계적 문헌고찰, 방법론적 질

*이 논문은 한국보건의료연구원에서 체계적 문헌고찰 방법론의 최신 업데이트를 위해 수행한 연구과제인 'NECA 비독립위험 평가도구 매뉴얼(2021)'에서 일부 발췌하였음을 밝힌다.

Address reprint requests to : Seo, Hyun-Ju

College of Nursing, Chungnam National University, 266 Munhwa-ro, Jung-gu, Daejeon 35015, Korea

Tel: +82-42-580-8320 Fax: +82-42-580-8309 E-mail: hjseo20@cnu.ac.kr

투고일: 2022.12.02 심사완료일: 2022.12.09 게재확정일: 2022.12.29

분석은 임상실무지침의 권고안을 내리는데 사용되는 근거의 타당성과 신뢰성을 훼손시켜 잠재적으로 잘못되고 신뢰할 수 없는 결과를 초래하여,⁵⁾ 근거의 위계상 이러한 유형의 연구가 갖고 있는 명성과 영향력을 고려할 때 사회적 관점에서 해로울 수 있다.⁴⁾

이로 인해 연구자나 보건의료전문가가 비평없이 단일 체계적 문헌고찰의 결과를 수용하는 것은 위험할 수 있다. 따라서, 이들은 출판된 체계적 문헌고찰 논문을 읽고 그 결론을 받아들이기 전에 해당 문헌의 방법론적 내적 타당도를 평가해야 하며, 이를 고려하여 결론을 해석하고 활용해야 한다. 체계적 문헌고찰의 방법론적 타당도를 검토하기 위한 비평적 평가도구로는 A Measurement Tool to Assess systematic Reviews (AMSTAR),⁶⁾ The Critical Appraisal Skills Programme (CASP) Checklist for systematic review,⁷⁾ The Scottish Intercollegiate Guidelines Network (SIGN) Methodology checklist: systematic reviews and meta-analyses,⁸⁾ The Joanna Briggs Institute (JBI) Critical Appraisal Checklist for systematic reviews and research syntheses,⁹⁾ NIH quality assessment tool for systematic reviews and meta-analyses,¹⁰⁾ The Risk of Bias in Systematic Review (ROBIS) tool¹¹⁾ 등이 접근가능하다.¹²⁾

이 중에서 최근에 개발되어, 중재에 대한 무작위대조연구와 비무작위연구들에 대한 체계적 문헌고찰의 질평가에 적합한 도구는 AMSTAR의 갱신판인 AMSTAR 2이며, 중재연구, 진단검사의 정확도, 임상적 예측, 원인 연구, 예후 연구들에 대한 체계적 문헌고찰의 비뚤림 위험 평가도구로는 ROBIS가 권장된다.¹³⁾ 최근 네트워크 메타분석은 연구 네트워크 전반에 걸쳐 직접 및 간접적 근거를 결합하여 단일 분석에서 동시에 세 개 이상의 중재를 비교하는 기법으로, 중재의 순위(ranking) 및 계층(hierarchy)을 추정할 수 있게 한다.¹⁴⁻¹⁵⁾ 네트워크 메타분석이 포함된 체계적 문헌고찰을 평가하기 위한 비뚤림 위험 평가도구는 현재 캐나다 연구그룹에서 CHIR(Canadian Institutes of Health Research) 연구비 지원을 받아 개발 중에 있다.¹⁶⁾

본 원고에서는 AMSTAR 2에 대한 평가항목에 대한 설명 및 평가한 체계적 문헌고찰 결과의 전반적인 신뢰도 평가 방법에 대해 간략히 소개하고자 한다.

AMSTAR 2 도구의 소개

2007년도에 개발되고 타당도가 검증된 AMSTAR는 무작위 대조연구에 대한 체계적 문헌고찰의 방법론적 질 평가도구로 가장 널리 사용되어왔지만, 연구수행의 전반적 측면을 평가도구 내에서 다루도록 수정될 필요가 있었다.¹⁷⁾ AMSTAR의 제한점은 비무작위 연구에 대한 체계적 문헌고찰의 비뚤림 위험 평가항

목을 포함하고 있지 않고, 평가항목에 대한 명확한 판단지침이 없어 평가자의 주관적 판단으로 인하여 평가결과와 신뢰성과 반박가능성을 떨어뜨릴 수 있다. 또한, 전반적 신뢰도 평가 기준이 제시되지 않아 체크리스트 도구임에도 불구하고 연구자들에 의해 종종 척도 도구(scoring system)처럼 해석되기도 하였다.¹⁸⁾

이러한 이유로 AMSTAR 개발그룹은 보건의료 중재에 관한 무작위대조연구 또는 비무작위연구에 대한 체계적 문헌고찰의 비평적 평가도구로 AMSTAR 2를 2017년에 출판하였다.

1. AMSTAR와 AMSTAR 2의 비교¹⁷⁾

AMSTAR 2는 AMSTAR의 11개 평가항목 중 10개 항목은 일부 용어만 바뀌어 그대로 존속되고 일부 항목이 새로 추가되어 총 16개의 평가항목으로 구성되어 있다. AMSTAR 2의 영역별 질문은 ‘예’ 대답이 긍정적인 결과를 나타내도록 구성되어 있고 평가 응답을 단순화하였다. AMSTAR의 ‘예’, ‘아니오’, ‘대답할 수 없다’, ‘적용할 수 없다’의 4가지 응답에서 ‘예’, ‘아니오’, ‘표준에 부분적 순응을 보인 경우’ ‘일부 예로 평가한다. 기존의 ‘대답할 수 없다’와 ‘적용할 수 없다’는 없어졌으며, AMSTAR 2는 관련 정보가 없다면 ‘아니오’로 평가해야 한다.

연구 선택과 자료 추출의 중복성을 두 문항으로 분리하였으며, 연구비 출처의 영향을 개별 연구 수준과 전체 문헌고찰 수준으로 구분하여 세분화시켰다. 일부 평가항목에서 무작위대조연구와 비무작위연구를 나누어서 비뚤림 위험을 평가하도록 하였다(예. 개별연구의 비뚤림 위험 평가 도구의 적절성, 메타분석 수행방법의 적절성). AMSTAR에서 포함기준에 출판상태(예: 회색 문헌) 사용여부에 대한 내용이 포괄적인 문헌검색 평가항목에 포함되었다. 새로이 네 개의 평가항목이 추가되었는데, 두 항목은(PICO [Population, Intervention, Comparison, Outcome]의 정교화, 근거합성 과정에서 비뚤림위험의 처리방법) ROBINS-I¹⁹⁾에서 왔으며, 두 항목은(이질성의 가능한 원인과 의미에 대한 고찰, 문헌고찰에 포함된 연구설계 선택의 정당화)은 비무작위 연구설계를 다루기 위하여 기존 AMSTAR 문항에서 일부 변형되었다.

AMSTAR 2는 평가자의 판단을 돕기 위한 사용자 지침(user guide)을 제공한다. 그러나 이 문서는 대부분 임상적 중재에 대한 체계적 문헌고찰이 수행된 경우 기준을 어떻게 적용해야 하는지에 대한 것이기 때문에, 개별 사용자에게 따라 개별 영역을 다룰 때 또는 전반적인 비뚤림 위험을 평가할 때 이 지침에서 벗어나는 것이 필요할 수도 있다. 그러한 경우에 지침에서의 이탈을 문서화하는 것이 필요하다.

2. AMSTAR 2 평가항목^{17,20)}

AMSTAR 2의 16개 평가항목은 실제 체계적 문헌고찰이 수행되는 과정 전반에 대한 방법론적 질을 평가하는 과정으로 구성되어 있다. 16개의 평가항목을 간략히 살펴보면 다음과 같다.

1) PICO의 정교화

관심 연구질문을 PICO(인구집단, 중재, 대조군, 중재결과) 형식으로 체계화하며, 이러한 PICO의 요소는 적격연구 선택을 위한 포함기준에 기술되어야 있어야 한다.

2) 리뷰 프로토콜

체계적 문헌고찰은 관찰연구의 측면을 갖고 있어서, 고찰을 시작하기 전에 고찰 방법론을 합의하여 문서화한 리뷰 프로토콜이나 가이드를 가지고 있거나 등록하였다는 기술이 있어야 한다.

3) 연구설계 선택의 정당화

체계적 문헌고찰에 포함되는 연구 유형이 자의적이면 안 된다. 이 영역은 다소 소홀히 다루어진 것도 사실이며, 평가 기준을 충족하는지에 대한 판단이 어려울 수 있다. 체계적 문헌고찰에 무작위대조연구만 또는 비무작위연구만 포함한 이유에 대한 설명, 또는 무작위대조연구와 비무작위연구 모두를 포함한 이유에 대한 설명이 있어야 한다.

4) 포괄적 문헌검색

체계적 문헌고찰에 있어 검색의 중요성은 잘 알려져 있으므로 기존 AMSTAR와 거의 유사하다. 모든 언어의 출판물을 찾아야 하며, 언어 제한이 있다면 정당화해야 한다. 경우에 따라 회색 문헌 검색의 필요성이 있는 때도 있다. 회색 문헌이 중요하게 여겨지는 경우, 저자들은 적절한 자료원을 검색해 보아야 하는데 임상시험 등록처(레지스트리), 학술대회 초록집, 학위 논문 등이 검색되어야 한다.

5) 문헌선정

최소한 두 명의 평가자가 독립적으로 적격연구를 선택한 후 최종 포함 여부를 합의한다. 또는 두 명의 평가자가 일부 표본에 대해 적격연구를 선정하였고, 좋은 일치도를 보여(예: kappa 통계량 0.8 이상), 나머지는 한 평가자가 문헌을 선정하는 때도 가능하다.

6) 자료 추출

자료 추출을 수행하는 독립적인 평가자가 적어도 두 명 이상

있어야 하나, 일부 표본에 대해 자료 추출에 대한 일치도가 kappa 통계량 0.8 이상을 달성하여 나머지 자료 추출을 한 명이 수행하는 것도 가능하다.

7) 배제연구 목록 및 배제사유 제시

전문까지 검토하였지만, 부적절한/부적합한 인구집단, 중재 및 비교 중재 등의 다양한 이유로 문헌고찰에서 배제된 연구에 대한 목록을 제시해야 한다. 유의해야 할 사항은 비뚤림위험에 근거하여 배제여부를 결정해서는 안 되며, 비뚤림위험은 별도로 문헌고찰 과정에서 다루어져야 한다.

8) 포함연구 세부 사항 제시

인구집단, 중재, 비교군, 중재결과, 연구설계, 연구세팅, 추적관찰 시점에 대한 상세한 설명이 제공되어야 한다. 또한, 중재효과의 이질성을 연구하기 위한 틀(예: 질병의 중증도, 약의 용량, 연령 범위, 임상 세팅 등)도 제공되어야 한다.

9) 개별연구의 비뚤림 위험 평가

문헌고찰 저자는 비뚤림 위험 평가에 체계적인 접근법을 사용해야 하며, 가능한 적절히 개발된 평가 도구를 사용해야 한다. 예를 들면, 저자들은 ROB 2,²¹⁾ ROBINS-I¹⁹⁾와 같은 코크란 비뚤림 위험평가도구를 기반으로 무작위 대조연구와 비무작위 연구에 대한 비뚤림 평가 영역을 고려해야 한다.

10) 연구비 출처 보고

상업적으로 후원되는 연구가 독립적으로 연구비를 지원받은 연구보다 후원자의 제품에 유리한 연구결과가 나올 가능성이 더 크다. 따라서, 문헌고찰 저자는 문헌고찰에 포함된 각 연구에 대한 연구비 출처를 상세히 기술하거나, 개별 연구논문에 정보가 제공되지 않았다는 것을 기술하는 것이 중요하다.

11) 메타분석이 수행된 경우, 통계적 결합의 적절성

문헌고찰 저자는 리뷰 프로토콜에 메타분석을 통해 자료를 결합하는 합당한 이유를 제시하여야 하며, 이러한 원칙에는 연구가 인구집단, 비교군, 중재의 측면에서 호환되는 정도와 단일 통합 효과의 가치(예: 호환 가능하지만 검정력이 낮은 여러 연구들)가 포함된다. 적절한 가중치법을 이용하여 연구결과를 결합했으며, 이질성이 있다면 그 원인을 조사해야 한다. 또한, 무작위 대조연구와 비무작위연구 모두가 고찰에 포함되었을 때, 무작위 대조연구와 비무작위연구의 요약 추정치를 별도로 보고하여야 한다. 이 항목은 메타분석을 보고 한 경우에만 평가하고, 메타분

석이 없으면 ‘메타분석 없음’에 체크 한다.

12) 메타분석이 수행된 경우, 개별연구의 비뚤림 위험이 합성결과에 미치는 영향 고려

이 영역은 새로운 평가항목으로, 문헌고찰 저자는 비뚤림 위험이 높다고 평가된 개별연구를 포함시키거나 제외시킴에 따라 결과가 어떻게 달라지는지에 대해 조사해야 한다. 문헌고찰 저자가 높은 질의 무작위대조연구들만 포함한 경우, 연구결과에 미치는 비뚤림의 잠재적 영향에 대한 논의가 거의 필요 없을 수 있다. 그러나, 다양한 질을 가진 무작위대조연구를 포함한 경우, 비뚤림 위험의 영향을 회귀분석으로 평가하거나, 비뚤림 위험이 낮은 연구만으로 통합효과크기를 추정해야 한다. 비무작위연구에서의 경우, 문헌고찰 저자들은 비뚤림위험이 낮거나 증등도인 연구, 또는 비뚤림위험이 낮은 연구만으로 통합효과크기를 추정해야 한다. 메타분석 또는 회귀분석과 같은 다른 자료합성 기술이 수행되지 않은 경우에도, 저자는 개별 연구결과에 대한 비뚤림 위험의 영향에 대한 의견을 제시해야 한다.

13) 고찰결과 해석시 개별연구의 비뚤림 위험 고려

비뚤림 위험 평가를 더욱 강조하기 위하여 기존 AMSTAR 평가항목을 수정한 항목이다. 고찰결과를 해석하고 논의하며 결론을 도출하거나 권고안을 제시할 때, 비뚤림 위험의 잠재적 영향을 명시적으로 언급해야 한다.

14) 이질성에 대한 설명 및 고찰

연구결과에서 유의미한 이질성이 없다면 문제가 되지 않는다. 그러나 메타분석 결과에서 이질성이 있는 경우, 이질성의 원인을 조사하고, 이질성이 문헌고찰 연구결과에 미치는 영향에 대해 논의해야 한다. 이질성의 주요 원인인 PICO 요소와 연구설계에 따른 비뚤림위험 영역을 조사하였는지가 중요하다.

15) 출판 비뚤림(소규모연구 비뚤림)에 대한 적절한 조사

출판 비뚤림에 대한 깔대기 그림(funnel plot) 또는 Egger's test와 같은 통계적 검정을 통해 평가하며, 그 결과가 유의하면 출판 비뚤림이 있다는 의미이다. 하지만, 검정이 민감하지 않기 때문에 검정결과가 유의하지 않은 것이 출판 비뚤림이 없음을 보장하지는 않는다. 깔대기 그림을 제시하기 위해서는 최소한 10개 이상의 연구가 필요하다. 중요한 것은 문헌고찰 저자가 1) 심층적이고 집중적인 문헌검색을 통해 출판 비뚤림을 조사하기 위해 최선을 다했는지, 2) 연구결과에 대한 해석 및 고찰에 출판 비뚤림의 영향을 고려하였는지, 3) 현재의 연구결과를 무효화하는데 필요한

누락된 ‘무효(null)’ 연구의 수를 결정하기 위해 민감도 분석(예, Trim and fill method)²²⁾을 수행하였는지 여부이다.

16) 문헌고찰 저자의 잠재적 이해 상충 보고

이해 상충이 있는 연구는 그렇지 않은 연구보다 증재에 유리한 연구결과를 도출할 수 있으며, 이는 체계적 문헌고찰도 마찬가지이다. 따라서 체계적 문헌고찰 저자도 직접적인 연구비 출처 및 이해 상충을 보고해야 한다. 이해 상충이 없다고 보고하거나 재정적 이해 상충이 있는 경우 이를 어떻게 관리하였는지를 보고해야 한다.

국문 AMSTAR 2 평가서식지는 Table 1과 같다. 국문 AMSTAR 2 도구는 한국보건의료연구원에서 개발한 『비뚤림위험평가 도구 매뉴얼』내의 Part 1. 체계적 문헌고찰 비뚤림위험 평가도구, 1장 AMSTAR 2(20~23페이지)를 참고하도록 한다.²⁰⁾

3. 체계적 문헌고찰 결과의 전반적인 신뢰도(confidence) 평가

체계적 문헌고찰 및 메타분석 수행의 모든 단계가 중요하지만, 개발그룹은 7개 영역이 고찰의 타당성과 결론에 결정적인 영향을 미칠 수 있다고 하였다(Table 2). 비뚤림 위험이 핵심적 영역들 중 2개(비뚤림 위험이 적절하게 평가되었는지 여부와 비뚤림 위험이 고찰 결과에 미치는 영향을 고려하였는지 여부)를 차지하였는데, 비뚤림 위험에 중요성을 부여한 이유는 AMSTAR 2가 비무작위연구를 포함하는 체계적 문헌고찰을 평가하는 데 사용되기 때문이다.

만약 질이 높은 무작위대조연구만을 포함한 경우에는 이것이 덜 중요할 수 있다. 임상시험협력그룹의 결과물을 기반으로 체계적 문헌고찰을 수행하는 경우와 같이 이미 알려진 문헌에 대한 합성을 할 때는 문헌검색(문항 4), 배제 문헌 목록 및 사유(문항 7), 출판 비뚤림 영역(문항 15)은 핵심적 영역으로 고려되지 않을 수 있다. 또한, 메타분석을 수행하지 않은 경우, 메타분석의 적절성 부분(문항 11)은 적용할 수 없는데 그런 경우에는 문헌고찰 저자가 연구결과를 요약할 때 비뚤림위험의 영향을 잘 고려하였는지를 평가해야 한다.

Table 2의 7개의 핵심적 영역은 개발자의 제안이며, 문헌고찰 평가자가 다른 중요한 문항을 추가하거나 대체할 수 있다. 예를 들어, 치료의 부작용에 관한 체계적 문헌고찰에서, 비무작위 연구를 포함하지 않았을 경우 문항 3(연구설계 선택의 정당화), 연구간 치료효과 크기의 상당한 이질성을 설명하지 못한 경우 문항 14(이질성에 대한 설명 및 고찰)가 핵심적 영역에서의 결함으로 평가될 수 있다.¹⁷⁾

Table 1. Korean Version of the AMSTAR 2 Evaluation Template

질문	판단
1 체계적 문헌고찰의 연구 질문과 포함기준에는 PICO의 구성요소가 포함되었는가?	☐ 예 ☐ 아니오
2 체계적 문헌고찰 방법론이 실제 문헌고찰을 시행하기 전에 확립되었으며, 보고서에는 프로토콜로부터 중대한 이탈이 있는 경우 이에 대한 정당화(합당한 이유)가 제시되었나?	☐ 예 ☐ 일부 예 ☐ 아니오
3 문헌고찰 저자는 문헌고찰에 포함될 연구설계 선택에 대해 설명하였나?	☐ 예 ☐ 아니오
4 문헌고찰 저자는 포괄적인 문헌 검색 전략을 사용하였는가?	☐ 예 ☐ 일부 예 ☐ 아니오
5 문헌고찰 저자는 연구 선택을 중복으로 수행하였는가?	☐ 예 ☐ 아니오
6 문헌고찰 저자는 자료추출을 중복으로 수행하였는가?	☐ 예 ☐ 아니오
7 문헌고찰 저자는 배제 연구에 대한 목록과 합당한 배제사유를 제공하였는가?	☐ 예 ☐ 일부 예 ☐ 아니오
8 문헌고찰저자는 포함된 연구들의 세부 사항을 적절히 기술하였는가?	☐ 예 ☐ 일부 예 ☐ 아니오
9 문헌고찰저자는 문헌고찰에 포함된 개별 연구의 비뚤림위험(ROB)을 평가하기 위해 만족스러운 도구를 사용하였는가?	☐ 예 ☐ 일부 예 ☐ 아니오 ☐ NRSI(RCT)만 포함
10 문헌고찰 저자는 문헌고찰에 포함된 연구들의 자금 출처에 대해 보고하였는가?	☐ 예 ☐ 아니오
11 메타분석을 수행하였다면, 문헌고찰 저자는 이에 대한 합당한 이유를 제시하였고, 연구결과와 통계학적 결합을 위해 적절한 방법을 사용하였는가?	☐ 예 ☐ 아니오 ☐ 메타분석 없음
12 메타분석을 수행하였다면, 문헌고찰 저자는 개별 연구의 비뚤림위험이 메타분석 연구결과나 다른 근거 합성에 미칠 잠재적 영향을 평가하였는가?	☐ 예 ☐ 아니오 ☐ 메타분석 없음
13 문헌고찰저자가 문헌고찰 결과를 해석/논의할 때 개별 연구의 비뚤림위험을 고려하였는가?	☐ 예 ☐ 아니오
14 문헌고찰저자는 문헌고찰 연구결과에서 발견된 이질성에 대해 만족스러운 설명을 하였는가?	☐ 예 ☐ 아니오
15 양적 합성을 하였다면, 문헌고찰저자는 출판비뚤림(소규모연구 비뚤림)에 대한 적절한 조사를 수행하고, 문헌고찰 결과에 미칠 수 있는 영향에 대해 고찰하였는가?	☐ 예 ☐ 아니오 ☐ 메타분석 없음
16 문헌고찰 저자는 문헌고찰 수행을 위한 자금지원을 포함하여 잠재적 이해상충에 대해 보고하였는가?	☐ 예 ☐ 아니오

*Reprinted from Kim et al.(2021)²⁰⁾ with permission of NECA.

Table 2. Critical domains of the AMSTAR 2

문항 2 : 문헌고찰 수행 전에 프로토콜 등록
문항 4 : 문헌검색의 적절성
문항 7 : 배제 문헌의 정당화
문항 9 : 포함 문헌에 대한 비뚤림위험 평가
문항 11 : 메타분석방법의 적절성
문항 13 : 문헌고찰의 연구결과 해석시 비뚤림위험의 고려
문항 15 : 출판비뚤림의 존재와 영향 평가

AMSTAR 2는 전체적인 '점수'를 매기게 되어 있지 않다. 높은 점수 자체로 부적절한 문헌 검색이나 체계적인 문헌고찰에 포함된 개별 연구의 비뚤림 위험을 적절히 평가하지 못한 것과 같은 특정 영역의 결정적인 약점이 감추어질 수도 있다. 오히려, 평가자는 각 영역에 대한 부적절한 평가가 고찰결과의 신뢰도에 미치는 잠재적 영향을 고려해야 한다. 왜냐하면, 그러한 결함은 체계적 문헌고찰에 대한 신뢰도를 크게 약화할 수 있기 때문이다. Figure 1은 핵심적 영역과 비핵심적 영역에서 확인된 결함을 가지고 체계적 문헌고찰 결과의 전반적 신뢰도를 해석하는 틀이다. 이것은 개발자가 제시한 조언으로 평가자는 검토 중인 고찰에서 가장 중요한 영역을 결정하여 체계적 문헌고찰 결과에 대한 신뢰도를 평가해야 한다.¹⁷⁾

맺음말

AMSTAR 2 개발그룹의 AMSATAR와 AMSTAR 2간의 평가 소요시간을 비교한 연구에서, AMSTAR 2의 평가 소요시간은 평균 15~32분으로 AMSTAR의 10-15분보다 길었다.¹⁷⁾ 이는 AMSTAR 2의 평가항목이 기존 11개에서 16개로 증가되었고,

비무작위연구를 포함한 체계적 문헌고찰을 평가할 때 어떤 교란 요인이 중요한지, 고려 중인 문헌고찰과 관련이 있는 선택 비뚤림과 측정 비뚤림을 확인하는 방법에 대한 방법론적 고려가 필요하기 때문일 것이다.

한편, 코크란그룹에서 체계적 문헌고찰연구의 비뚤림 위험 평가도구로 개발된 ROBIS와 AMSTAR 2간의 평가 소요시간을 비교한 연구에서도 AMSTAR 2가 ROBIS보다 약간 평가 소요시간이 길었다(AMSTAR 2: 평균 18분, ROBIS: 평균 16분).¹³⁾ 그러나 이는 영어권 평가자의 소요시간이었기 때문에 우리나라의 상황과는 다를 수 있다.

또한, AMSTAR 2의 평가자간 일치도에서, AMSTAR 2는 핵심적 영역에 대한 문항들에서 다른 문항들에 비해 일치도가 좋은 것으로 나타났다. 문항 2(리뷰 프로토콜), 문항 7(배제연구 목록 및 배제사유 제시), 문항 10(연구비 출처 보고)은 거의 완벽하게 일치하였으며, 문항 14(이질성에 대한 설명 및 고찰)를 제외한 모든 문항에서 중간 정도의 일치도를 보였다. AMSTAR 2의 평가자간 전반적인 일치도는 Gwet's AC 통계량 중앙값 0.65로 상당한 일치도를 보여 체계적 문헌고찰 비평적 평가도구로 유용성이 확인되었다.¹³⁾

근거기반 임상실무 및 정책결정에 최상의 이용 가능한 정보를 제공하기 위함이라는 체계적 문헌고찰의 본래의 목적에 기여하기 위하여, 연구자는 AMSTAR 2를 적용하여 비평적 평가를 수행할 때 변호사가 아닌 판사와 배심원의 사고방식을 채택해야 한다. 즉, 판사와 배심원단은 가장 공정한 판단을 내리기 위해 증거를 회의적으로 평가하는 반면, 변호사의 접근 방식은 주장의 한 측면에 대해 최선의 판례를 만드는 것이기 때문이다.²³⁾ 이러한 사고방식은 체계적 문헌고찰을 수행하는 연구자 뿐만 아니라, 체계적 문헌고찰을 검색하여 읽고 해석하여 임상적 의사결정에 활

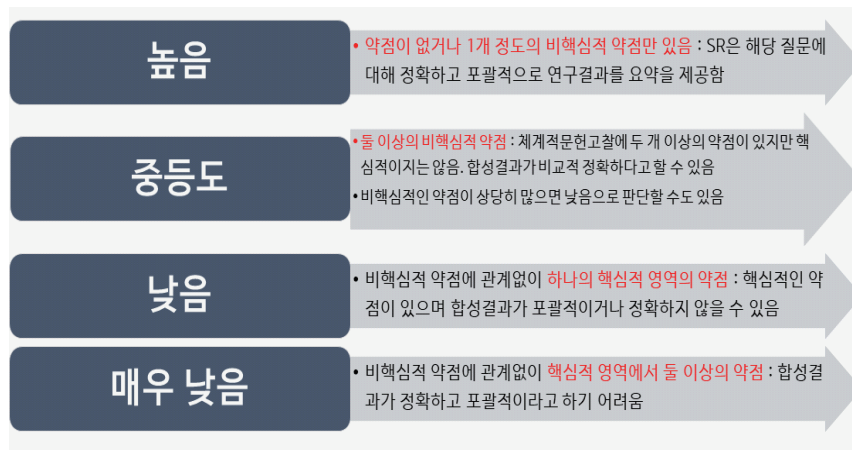


Figure 1. Rating overall confidence in the results of the systematic review.

용하고자 하는 보건의료전문가에게도 필요한 접근법일 것이다.

역학에 대한 심화교육을 반드시 받지 않아도 사용할 수 있어 실용적인 비평적 평가도구¹⁷⁾로 알려진 AMSTAR의 갱신판인 AMSTAR 2는 평가자간 일치율이 높고, 평가항목이 직관적이며, 평가지침이 포함되어 있어 체계적 문헌고찰 연구결과에 대한 신뢰도를 평가하고자 할 때 적용 가능한 비평적 평가도구로 판단된다. 그러나 AMSTAR 2는 체계적문헌고찰을 수행하는 연구자들을 위한 간단한 체크리스트나 편리한 교육지원 역할을 할 수는 있지만, 체계적 문헌고찰 수행의 방법론적 엄격성과 접근법을 상세히 설명해주지는 못하는 제한점을 가지고 있다. 따라서, 체계적 문헌고찰 연구수행에 대한 포괄적 조언이 필요한 연구자는 중재에 대한 체계적문헌고찰을 위한 Cochrane 핸드북,²⁴⁾ NECA의 의료기술평가방법론-체계적 문헌고찰¹⁵⁾ 및 JBI의 체계적 문헌고찰 매뉴얼²⁵⁾ 등의 방법론 지침서를 활용해야 할 것이다.

ORCID

서현주 0000-0001-9019-1135

REFERECES

- Mulrow CD. Rationale for systematic reviews. *BMJ* 1994;309:597-9. doi:10.1136/bmj.309.6954.597
- Al-Ghimlas F. The philosophy of evidence-based clinical practice: Is evidence enough? *Ann Thorac Med*. 2013;8(3):131-2. doi: 10.4103/1817-1737.114282
- Aromataris E, Pearson A. The systematic review: an overview. *Am J Nurs*. 2014;114(3):53-8
- Ioannidis JP. The mass production of redundant, misleading, and conflicted systematic reviews and meta-analyses. *Milbank Q*. 2016;94(3):485-514. doi: 10.1111/1468-0009.12210
- Lunny C, Ramasubbu C, Puil L, Liu T, Gerrish S, Salzwedel DM, et al. Over half of clinical practice guidelines use non-systematic methods to inform recommendations: A methods study. *PLoS One*. 2021;16(4):e0250356. doi: 10.1371/journal.pone.0250356
- AMSTAR. AMSTAR checklist. available at: <http://www.amstar.ca/> [accessed on 26 Dec 2022]
- CASP. CASP checklist. available at: <https://casp-uk.net/casp-tools-checklists/> [accessed on 26 Dec 2022]
- Healthcare Improvement Scotland. SIGN checklist. available at: <https://www.sign.ac.uk/what-we-do/methodology/checklists/> [accessed on 26 Dec 2022]
- The Joanna Briggs Institute. Checklist for systematic reviews and research syntheses. available at: https://jbi.global/sites/default/files/2019-05/JBI_Critical_Appraisal_Checklist_for_Systematic_Reviews2017_0.pdf [accessed on 26 Dec 2022]
- NIH. Study quality assessment tools. available at: <https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools> [accessed on 26 Dec 2022]
- University of BRISTOL. ROBIS tool. available at: <http://www.bristol.ac.uk/population-health-sciences/projects/robis/robis-tool/> [accessed on 26 Dec 2022]
- Ma LL, Wang YY, Yang ZH, Huang D, Weng H, Zeng XT. Methodological quality(risk of bias) assessment tools for primary and secondary medical studies: what are they and which is better? *Mil Med Res*. 2020;7(1):7. doi: 10.1186/s40779-020-00238-8
- Perry R, Whitmarsh A, Leach V, Davies P. A comparison of two assessment tools used in overviews of systematic reviews: ROBIS versus AMSTAR-2. *Syst Rev*. 2021;10(1):273. doi: 10.1186/s13643-021-01819-x
- Anna C, Deborah MC, Tianjing L, Julian PTH, Georgia S. Chapter 11: Undertaking network meta-analyses. *Cochrane handbook for systematic reviews of interventions*. version 6.3, 2022. <https://training.cochrane.org/handbook/current/chapter-11>
- Kim SY, Park DA, Seo HJ, Shin SS, Lee SJ, Lee M, et al. Health technology assessment methodology: Systematic review, National Evidence-based Healthcare Collaborating Agency, Seoul, 2020.
- Lunny C, Tricco AC, Veroniki AA, Dias S, Hutton B, Salanti G, et al. Methodological review to develop a list of bias items used to assess reviews incorporating network meta-analysis: protocol and rationale. *BMJ Open*. 2021;11(6):e045987. doi: 10.1136/bmjopen-2020-045987
- Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ*. 2017;358:j4008. doi: 10.1136/bmj.j4008.
- Faggion CM Jr. Critical appraisal of AMSTAR: challenges, limitations, and potential solutions from the perspective of an assessor. *BMC Med Res Methodol*. 2015 Aug 13;15:63. doi: 10.1186/s12874-015-0062-6
- Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ*. 2016;355:i4919. doi: 10.1136/bmj.i4919
- Kim SY, Park DA, Seo HJ, Shin SS, Lee SJ, Jang BH, et al. NECA's guidance for assessing tools of risk of bias, National Evidence-based Healthcare Collaborating Agency, Seoul, 2021.
- Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ*. 2019;366:l4898. doi: 10.1136/bmj.l4898
- Shim SR, Shin IS, Bae JM. Intervention meta-analysis using STATA software. *J Health Info Stat*. 2016;41(1):123-134. doi: <https://doi.org/10.21032/jhis.2016.41.1.123>
- Baumeister RF. Writing a literature review. In: MJ Prinstein, MD Patterson, editors. *The portable mentor: Expert guide to a successful career in psychology*, 2nd ed. New York: Springer. 2013. p. 119-132.
- Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. editors. *Cochrane handbook for systematic reviews of interventions* version 6.3(updated February 2022). Cochrane, 2022. Available at: <http://www.training.cochrane.org/handbook>
- Aromataris E, Munn Z. Chapter 1: JBI systematic reviews. In: Aromataris E, Munn Z, editors. *JBI manual for evidence synthesis*. JBI, 2020. Available at: <https://synthesismanual.jbi.global>. <https://doi.org/10.46658/JBIMES-20-02>