의료기술평가방법론

체계적 문헌고찰



Copyright © 한국보건의료연구원, 2020. All rights reserved. First edition.

본 책자는 공공 소유로서 비영리 목적으로의 인용, 배포 및 복제는 한국보건의료연구원의 별도의 승인 없이도 가능합니다. 그러나 특정 그림이나 표의 경우 원저작권자의 승인 없이는 사용할 수 없습니다.

의료기술평가방법론 : 체계적 문헌고찰

인쇄: 2020년 5월 31일 초판 1쇄 발행
발행인: 한광협
발행처: 한국보건의료연구원
서울특별시 중구 퇴계로 173 남산스퀘어빌딩 7층
Tel. 02-2174-2700
Homepage: www.neca.re.kr

ISBN 978-89-6834-643-9
ISBN 978-89-6834-003-1 (세트)

Copyright © National Evidence-based Healthcare Collaborating Agency, 2020.
First edition.

Published by NECA
May 2020

All rights reserved. Reproduction of this book by photocopying or electronic means for non-commercial purposes is permitted except those copyrighted materials noted for which further reproduction is prohibited without the specific permission of copyright holders. Otherwise, no part of this book may be reproduced, adapted, stored in a retrieval system or transmitted by any means, electronic, mechanical, photocopying, or otherwise without the prior written permission of NECA.

Comments and suggestions on this book can be made at neca@neca.re.kr



집필진

본 매뉴얼은 다음과 같이 인용될 수 있습니다.

김수영, 박동아. 서현주, 신승수, 이수정, 이민, 장보형, 차영주, 최인순, 박균익.
의료기술평가방법론:체계적 문헌고찰. 한국보건의료연구원. 2020.

Suggested citation:

SY Kim, DA Park, HJ Seo, SS Shin, SJ Lee, M Lee, BH Jang, YJ Cha, IS Choi, KI Park.
Health Technology Assessment Methodology:Systematic Review, National Evidence-based Healthcare Collaborating Agency, Seoul, 2020.

김수영 교수
한림대학교 의과대학 가정의학교실

박동아 연구위원
한국보건의료연구원

서현주 부교수
충남대학교 간호대학 간호학과

신승수 교수
아주대학교 의과대학 호흡기내과

이수정 연구강사
연세대학교 원주의과대학 간호학과

이 민 부연구위원
한국보건의료연구원

장보형 부교수
경희대학교 한의과대학 예방의학교실

차영주 대리
코반스코리아 임상시험사업본부

최인순 연구위원
한국보건의료연구원

박균익 연구원
한국보건의료연구원



발간사

근거기반 정책 의사결정의 확산을 기대하며…

근거기반 의사결정을 지원하기 위하여 근거기반 보건의료, 근거기반의학 등의 개념이 발달하였고 이를 실현하는 접근법으로 주요하게 활용되는 방법론이 체계적 문헌고찰입니다. 근거기반의 개념과 방법론이 정립된 초기에는 의료기술의 안전성 및 유효성을 평가하기 위하여 정통의 체계적 문헌고찰을 반드시 수행해야 한다고 인식하였으나 이제는 기존의 체계적 문헌고찰의 출간도 증가하고 보다 신속하게 평가를 요구하는 사례도 많이 대두되었습니다.

또한 체계적 문헌고찰의 세부적인 방법론이 꾸준히 개발되고 발전하게 되었으며 이는 최근 중재법 기반의 체계적 문헌고찰에 대한 국제적인 지침서인 코크란 핸드북의 제 6판이 출간되는 계기였다고 생각합니다.

그동안 한국보건의료연구원에서는 보건의료정책 의사결정을 지원하기 위하여 기존 의료기술에 대한 안전성 및 유효성, 경제성 등을 평가하는데 활용되는 체계적 문헌고찰 방법론을 국제수준에 적합하게 수행하기 위하여 2011년 『NECA 체계적 문헌고찰 매뉴얼』을 발간한 바 있습니다. 그러나, 매뉴얼이 발간된지 10년 가까이 지나면서 국제적으로도 체계적 문헌고찰의 방법론은 다양화되고 여러 활용되는 기준과 도구가 발전하였습니다.

따라서, 연구원뿐 아니라 국내 연구자들이 활용할 수 있도록 코크란 핸드북의 제6판(2019)의 실제적인 방법론에 대한 안내 및 최신 방법론적 동향을 반영한 종합적인 지침서가 필요한 상황이라고 판단하였습니다.

본 NECA 체계적 문헌고찰 저서에서는 프로토콜 작성에서 결론도출 및 보고서 작성에 이르는 체계적 문헌고찰 수행 단계별 내용뿐 아니라 체계적 문헌고찰을 수행하면서 고려하여야 하는 다양한 사항들에 대해 심도있게 다루고 있습니다. 여기에는 형평성, 중재 복합성, 개별 환자자료 분석, 기존 체계적 문헌고찰의 활용, 신속 체계적 문헌고찰 방법, 경제성 분석 근거, 질적평가 연구, 유해 효과 및 전향적 접근에 대한 이슈 사항 등을 정리하였습니다.

적절한 체계적 문헌고찰을 바탕으로 한 의료기술평가보고서는 근거기반 보건의료의 시금석이라고 할 수 있습니다. 근거에 기반한 타당하고 신뢰할 수 있는 정보 제공은 의사 및 보건의료 정책가 등 전문가들이 진료 또는 보건의료정책 수립 중 내리는 의학적 결정의 확신성을 높이고 궁극적으로 보건의료의 질을 향상시키는 결과를 가져올 수 있을 것이라 기대합니다.

한국보건의료연구원의 이번 결과물을 계기로 신뢰도 높은 연구 성과물 및 의료기술평가 보고물의 개발과 이에 대한 현장에서의 활용이 더욱 활성화되고 연구방법론의 지속적인 발전을 앞장서서 선도하기를 바랍니다.

2020년 5월

한국보건의료연구원장

서문

체계적 문헌고찰이라는 방법론은 근거중심의학의 발전과 함께 눈부신 발전을 거듭해왔습니다. 근거는 선택하는 것이 아니고, 종합하는 것이라는 근거중심의학의 핵심 개념을 체계적 문헌고찰이 구체화하였기 때문입니다. 체계적 문헌고찰 방법론은 코크란 연합 등의 활동으로 구체화 되었으며 현재까지 수많은 의학적 의사결정에 기여하였습니다.

최근 들어 체계적 문헌고찰의 양적 증가는 매우 급격하다고 볼 수 있습니다. 일례로 PubMed에서 출판형태(publication type)로 메타분석(meta-analysis)를 검색해보면, 체계적 문헌고찰의 수는 2000년 816건에서 2019년은 8742건으로 10배가 증가하였으며, 1990년에 273건에 비해서는 40배가 증가하였습니다. 하지만 이러한 양적 팽창이 반드시 바람직한 것만은 아닙니다.

심지어 질이 높지 않은 체계적 문헌고찰이 너무 많이 증가하여 잘못된 의사결정을 유도하기도 하고, 올바른 의사결정을 오히려 방해한다는 의견도 있을 정도입니다.

이러한 현상에 대해 코크란을 중심으로 한 방법론 그룹에서는 좀 더 엄격한 방법론을 체계적 문헌고찰에 도입하려는 움직임이 활발해지고 있으며, 체계적 문헌고찰의 등록(registration, PROSPERO, www.crd.york.ac.uk/prospero/), 체계적 문헌고찰 평가 도구의 개발(AMSTAR II, ROBIS), 체계적 문헌고찰의 표준 정립(MECIR, Methodological Expectations of Cochrane Intervention Reviews, https://methods.cochrane.org/methodological-expectations-cochrane-intervention-reviews) 등의 움직임으로 이어졌습니다.

최종적으로 2019년 6월 체계적 문헌고찰의 새로운 핸드북이 발표 되었습니다(https://training.cochrane.org/handbook). 새로운 코크란 방법론은 체계적 문헌고찰의 질을 높이기 위한 여러 가지 방법론은 제안하고 있습니다.

첫째, 연구결과 결측으로 인한 비뚤림위험에 대한 내용입니다. 과거 상당수 체계적 문헌고찰의 내용은 출판된 문헌에만 기초하여 편향된 연구결과를 도출할 수 있다는 깨달음입니다. 코크란 연합은 극복하기 위해 검색(임상연구 보고서와 규제 문서의 검색), 분석기법(개시 코호트의 제안), 비뚤림위험평가(비보고 비뚤림위험 평가) 등의 방법론을 제시하고 있습니다.

둘째, 비무작위 연구에 대한 내용입니다. 전통적으로 코크란 리뷰는 무작위배정 비교임상시험만을 포함하도록 권고하였지만, 다양한 부분에서 비무작위연구에 대한 문헌고찰이 필요해졌고, 비무작위 연구에 대한 체계적 문헌고찰 과정에서 필연적으로 생기는 비뚤림의 문제가 발생합니다. 코크란 연합은 이를 극복하기 위해서 중재복합성의 개념, 로직모델이라는 분석 틀의 도입, ROBINS-I라는 비무작위연구 비뚤림위험 평가 도구의 도입을 권고하였습니다.

셋째, 자료합성과 관련해서 여러 가지 새로운 방법론이 도입되었습니다. 과거 special issue로 다루어지던 네트워크 메타분석이 제 1부에서 다루어지면서 그 내용이 대폭 정비되었으며, 메타분석을 할 수 없을 때 자료를 효과적이고 과학적으로 제시할 수 있는 방법론들이 소개되었습니다.

이러한 내용뿐만 아니라 체계적 문헌고찰에서 이해상충을 어떻게 다룰 것인가에 대한 틀을 제시하였고, 체계적 문헌고찰 여러 과정의 시간을 단축하고 신뢰성을 높일 수 있는 반자동 소프트웨어에 대한 내용 등이 새롭게 도입된 내용이라고 할 수 있습니다.

지금까지 국내에서 체계적 문헌고찰에 대한 참고문헌은 한국보건의료연구원에서 2011년 발행된 "NECA 체계적 문헌고찰 매뉴얼"이 거의 유일한 것이었습니다. 체계적 문헌고찰에 대한 교육 프로그램은 한국보건의료연구원뿐만 아니라 한국 코크란 지회, 대한 의학회, 개별 의학 학회, 의과대학 등에서 이루어졌지만 새롭게 정립된 방법론의 내용을 쉽게 접할 수는 없었습니다.

애초에 이 책은 2014년에 나온 매뉴얼을 업데이트하는 것이 목표였습니다. 하지만 새로 발행된 코크란 핸드북이 방법론 변화의 폭이 커서 매뉴얼의 형태로 표준적 방법론을 제시하는 것이 힘들다고 판단하였고, 새로운 방법론으로 진행된 체계적 문헌고찰이 거의 없어서 매뉴얼에 꼭 필요한 내용인 예시와 팁 등을 제공하는 것이 힘들어졌습니다. 그런 이유로 현재와 같은 형태의 새로운 방법론을 소개하는 형태의 책을 발간하기로 하였습니다.

이 책에 있는 내용은 다음과 같은 내용을 담고 있습니다.

첫째, 전체적인 내용은 코크란 핸드북 6.0의 주요 내용을 소개하는 것입니다. 모든 내용을 다 담는 것은 힘들지만 전체적인 통일성을 잃지 않는 범위에서 요약과 핵심 내용을 제공하는 것을 목표로 하였습니다. 그리고 각 장의 말미에 MECIR를 소개하여 표준적인 내용을 제시하였습니다.

둘째, 코크란 핸드북에 없는 추가적인 내용을 일부 소개하였습니다. 그러한 내용에는 신속 체계적 문헌고찰, 기존의 체계적 문헌고찰의 활용이 있습니다.

셋째, 코크란 핸드북을 소개하면서 힘들었던 것 중의 하나는 용어의 번역에 대한 것입니다. 현재 체계적 문헌고찰 관련 용어는 아직 표준화되지 못한 것들이 많기 때문에 한국보건의료 연구원 연구진과 함께 용어 번역에 대해 논의하였고 합의에 이른 내용을 부록에 소개하였습니다.

이 책의 목적은 NECA에서 체계적 문헌고찰을 수행할 때 참조할 수 있는 방법론적인 내용을 소개하는 것입니다. 하지만 한국보건의료연구원 이외에서 체계적 문헌고찰 연구를 시행하는 연구자도 참고할 수 있습니다.

향후 이 책의 내용은 체계적 문헌고찰의 경험치가 쌓이면서 정제되고 정리되어 새로운 형태의 핸드북이나 매뉴얼로 발전할 수 있을 것입니다. 그렇게 하도록 해당 내용을 꼼꼼히 살펴 보고 적극적으로 활용하려는 노력이 필요할 것으로 생각합니다.

이 책을 만드는데 많은 수고를 하신 한국보건의료연구원의 박동아 연구위원, 최인순 연구위원, 이민 부연구위원, 아주대학교 신승수 교수, 경희대학교 장보형교수, 충남대학교 서현주교수, 연세대학교 원주의과대학 이수정교수, 코반스코리아 차영주님께 감사를 드립니다, 아울러 이 책의 내용에 대해 꼼꼼히 자문해 주신 가톨릭의과대학 임현우교수님, 안동대학교 장은진교수님, 고려대학교 김현정교수님, 중앙대학교 정선영교수님께도 감사의 마음을 전합니다. 끝으로 이 책 발행에 물심양면 지원을 아끼지 않으신 한광협 원장님께도 깊은 감사를 드립니다.

2020년 5월

김 수 영

목차

PART I . 체계적 문헌고찰 방법론

발간사

서문

제1장 연구 기획과 프로토콜 작성 ····· 1

제2장 주제범위 결정 ····· 9

제3장 선정기준결정과 PICO 범위 설정하기 ····· 23

제4장 문헌검색 ····· 38

제5장 자료 수집 ····· 60

제6장 비뚤림위험 평가의 개념과 무작위배정 임상연구 평가 ····· 80

제7장 비무작위 연구 포함 결정과 비뚤림위험 평가 ····· 97

제8장 비보고 비뚤림 평가 ····· 114

제9장 효과 크기 선택 ····· 127

제10장 메타분석의 준비 ····· 140

제11장 메타분석 수행 ····· 146

제12장 메타분석 할 수 없는 자료의 제시 ····· 175

제13장 네트워크 메타분석 ····· 180

제14장 자료분석과 결과제시 ····· 191

제15장 근거 수준 평가와 결론의 도출 ····· 207

PART II. 체계적 문헌고찰 방법론의 주요 이슈

제16장 형평성 ····· 227
제17장 중재의 복합성 ····· 235
제18장 개별 연구대상자 자료 분석 ····· 247
제19장 기존의 체계적 문헌고찰의 활용(Using Existing Systematic Reviews) ··· 257
제20장 신속 체계적 문헌고찰 ····· 265
제21장 경제성 근거 ····· 278
제22장 질적 근거 합성 ····· 289
제23장 특수 무작위배정 임상시험 포함 ····· 302
제24장 유해효과 ····· 313
제25장 근거 축적을 위한 전향적 접근 ····· 321
제26장 문헌고찰 Overview ····· 331
제27장 환자보고 중재결과(Patient-reported outcomes) ····· 349

부록. 관련용어 번역 ····· 359

PART I

체계적 문헌고찰 방법론



목차 PART I. 체계적 문헌고찰 방법론

제1장 연구 기획과 프로토콜 작성

1.1 체계적 문헌고찰을 하는 이유 ····· 1
1.2 핵심 질문 ····· 2
1.3 체계적 문헌고찰의 수행자 ····· 2
1.4 신뢰의 중요성 ····· 4
1.5 프로토콜 개발 ····· 4
1.6 데이터 관리 및 품질 보증 ····· 6
1.7 연구 기획 관련 MECIR ····· 7
1.8 참고문헌과 읽을거리 ····· 8

제2장 주제범위 결정

2.1 잘 만들어진 질문의 필요성 ····· 9
2.2 중재연구 문헌고찰의 목표 ····· 10
2.3 문헌고찰 질문의 범위를 정의하기 ····· 11
2.4 문헌고찰이 올바른 질문을 하고 있는지를 보장하기 ····· 15
2.5 문헌고찰의 구조화를 위한 방법 및 도구 ····· 17
2.6 주제 범위 관련 MECIR ····· 20
2.7 참고문헌과 읽을거리 ····· 21

제3장 선정기준결정과 PICO 범위 설정하기

3.1 서론 ····· 23
3.2 문헌고찰과 비교 PICO 명확화하기 ····· 24
3.3 선정 기준 관련 MECIR ····· 34
3.4 참고문헌과 읽을거리 ····· 36

제4장 문헌검색

4.1 개요 ····· 38
4.2 일반적인 문제들 ····· 39
4.3 검색원 ····· 40
4.4 검색 전략의 설계 ····· 45
4.5 검색 과정 문서화 및 보고 ····· 50
4.6 연구의 선택 ····· 50
4.7 문헌 검색 관련 MECIR ····· 55
4.8 참고문헌과 읽을거리 ····· 59

제5장 자료 수집

5.1 서론 60
5.2 자료 출처 60
5.3 수집 자료 63
5.4 자료 수집 도구 70
5.5 보고서에서 자료 추출 73
5.6 연구결과 추출 및 형식 변환 76
5.7 자료관리 및 공유 76
5.8 자료 수집 관련 MECIR 77
5.9 참고문헌과 읽을거리 79

제6장 비뚤림위험 평가의 개념과 무작위배정 임상연구 평가

6.1 비뚤림위험 평가의 개념 80
6.2 무작위배정 임상시험 비뚤림위험 평가 85
6.3 비뚤림위험 평가 관련 MECIR 95
6.4 참고문헌과 읽을거리 96

제7장 비무작위 연구 포함 결정과 비뚤림위험 평가

7.1 체계적 문헌고찰에 비무작위 연구를 포함 97
7.2 비무작위 연구에서 비뚤림위험 평가 102
7.3 참고문헌과 읽을거리 112

제8장 비보고 비뚤림 평가

8.1 개요 114
8.2 연구결과 결측으로 인한 비뚤림위험 최소화 116
8.3 합성과정에서 연구결과 결측으로 인한 비뚤림위험 평가를 위한 틀 118
8.4 비보고 비뚤림 관련 MECIR 125
8.5 참고문헌과 읽을거리 125

제9장 효과 크기 선택

9.1 자료 유형 및 효과 크기 127
9.2 효과 크기의 직접 추출 128
9.3 이분형 자료 129
9.4 연속형 중재결과 자료 131
9.5 순서형 중재결과 자료 136
9.6 계수자료와 비율 자료 137
9.7 사건발생 시간(time to event) 자료 138
9.8 참고문헌과 읽을거리 139

목차 PART I. 체계적 문헌고찰 방법론

제10장 메타분석의 준비

10.1 개요 ····· 140
10.2 합성의 일반적인 단계 ····· 141
10.3 메타분석의 예비 단계 ····· 142
10.4 합성 전 자료 확인 ····· 144
10.5 합성의 유형 ····· 144
10.6 참고문헌과 읽을거리 ····· 145

제11장 메타분석 수행

11.1 메타분석 소개 ····· 146
11.2 메타분석에 대한 일반적인 역분산 접근법 ····· 147
11.3 이분법 결과의 메타분석 ····· 151
11.4 연속형 중재결과의 메타분석 ····· 154
11.5 이분법적 결과 및 연속적 결과 결합 ····· 157
11.6 서수 결과 및 측정 척도의 메타분석 ····· 158
11.7 계수와 비율의 메타분석 ····· 158
11.8 사건발생 시간 자료(time-to-event)의 메타분석 ····· 159
11.9 이질성 ····· 161
11.10 이질성에 대한 조사 ····· 165
11.11 결측값 ····· 169
11.12 메타분석에 대한 베이지안적 접근법 ····· 171
11.13 민감도 분석 ····· 171
11.14 메타분석 수행 관련 MECIR ····· 173
11.15 참고문헌과 읽을거리 ····· 174

제12장 메타분석 할 수 없는 자료의 제시

12.1 효과 크기의 메타분석이 불가능한 이유 ····· 175
12.2 효과 크기의 메타분석이 불가능한 경우 통계적 합성 ····· 175
12.3 자료의 시각적 표시 및 표현 ····· 177
12.4 참고문헌과 읽을거리 ····· 179

제13장 네트워크 메타분석

13.1 네트워크 메타분석이란? ····· 180
13.2 네트워크 메타분석을 위한 중요개념 ····· 181
13.3 결과의 합성 ····· 184
13.4 네트워크 메타분석 결과에 대한 신뢰도 평가 ····· 187
13.5 네트워크 메타분석 제시 ····· 188

13.6 결론 ····· 190
13.7 참고문헌과 읽을거리 ····· 190

제14장 자료분석과 결과제시

14.1 자료 분석 ····· 191
14.2 결과제시 ····· 197
14.3 자료분석과 결과제시 관련 MECIR ····· 205
14.4 참고문헌과 읽을거리 ····· 205

제15장 근거 수준 평가와 결론의 도출

15.1 근거 수준 평가 ····· 207
15.2 결론 도출 ····· 214
15.3 근거 수준 평가와 결론 도출 관련 MECIR ····· 219
15.4 참고문헌과 읽을거리 ····· 219

목차 PART I. 체계적 문헌고찰 방법론

표 목차

표 2-1. 광범위 문헌고찰과 좁은 문헌고찰의 장단점 ····· 13
표 4-1. 검색의 민감도 및 정밀도 ····· 46
표 5-1. 체계적 문헌고찰의 정보 유형 ····· 64
표 6-1. ROB2에 포함된 비뚤림 영역과 다루는 이슈 ····· 89
표 6-2. 특정 결과에 대한 전반적인 비뚤림위험 평가 방법 ····· 91
표 7-1. 연구 특성 체크리스트. 각 항목에 대한 응답은 예, 아니오 또는 불확실로 기록한다(Reeves et al 2017) ····· 100
표 7-2. ROBINS-I 도구에 포함된 비뚤림 영역 ····· 106
표 7-3. 개별 비뚤림 영역에 대한 비뚤림위험 판단 ····· 108
표 7-4. 특정 중재결과에서 전반적 비뚤림위험의 판단 ····· 108
표 7-5. 추적관찰 연구에 대한 ROBINS-I tool ····· 109
표 7-6. 비통제 전후 연구에 대한 ROBINS-I tool ····· 111
표 7-7. 비교 전후 연구에 대한 ROBINS-I tool ····· 112
표 8-1. 다양한 비보고 비뚤림과 해당 유형의 정의 ····· 115
표 9-1. 메타분석의 주요 자료 유형 ····· 127
표 9-2. 임상연구에서 중재군과 대조군의 연구 결과(이분형 자료) ····· 130
표 9-3. 중재군과 대조군의 연구 결과(연속형 자료) ····· 131
표 9-4. 차이에 대한 표준편차(SD)가 없는 경우 ····· 134
표 9-5. 평균과 표준편차가 보고된 자료($평균 \pm SD$) ····· 134
표 9-6. 차이의 표준편차가 보고 되지 않은 자료($평균 \pm SD$) ····· 136
표 10-1. 결과 합성의 단계 ····· 141
표 10-2 요약 및 합성에 사용할 수 있는 방법 ····· 145
표 11-1. 중재군과 대조군의 연구결과(이분형 자료) ····· 151
표 11-2. 이질성의 종류 ····· 161
표 11-3. 메타분석에서 결측자료의 형태 ····· 170
표 15-1. 근거 수준 기술의 틀과 하향 평가 또는 상향평가에 대한 정당화 ····· 213
표 15-2. 개별 GRADE 도메인에서 제안한 연구 및 임상진료에 대한 영향 ····· 215
표 15-3. 결론을 위한 제안된 서술문 ····· 217

그림 목차

그림 6-1. 숲그림으로 비뚤림위험 제시의 예 ····· 83
그림 8-1. 연구자료 결측으로 인해 깔때기 그림에 비대칭이 생긴 경우····· 123
그림 8-2. 등고선 강화 깔때기 그림(contour-enhanced funnel plots). ····· 124
그림 13-1. 네트워크메타분석의 비교 ····· 181
그림 13-2. 간접비교 ····· 182
그림 13-3. 네트워크 도표 ····· 188
그림 14-1. '결과 요약표' 예 ····· 199
그림 14-2. 대안적인 '결과 요약표'예 ····· 200
그림 15-1. GRADE 접근의 근거 수준 ····· 209
그림 15-2. 중재결과에 따른 간접성에 대한 판단
(GRADEpro GDT에서 이용가능함) ····· 211

박스 목차

Box 8-1. 주요 우울증에서 미출판 자료 포함 여부에 따른 reboxetine의 효과 117
Box 8-2. 규제 문서에 대한 정보 ····· 117
Box 8-3. fibromyalgia에서 SSRI 효과를 위약과 비교하는 코크란 리뷰 :
세 개의 합성결과에 대한 연구결과의 가용성 메트릭스의 예 ····· 121
Box 8-4. 우울증 점수 메타분석에 사용할 수 있는 결과와 누락된 결과를
표시하는 숲 그림 ····· 122

제1장 연구 기획과 프로토콜 작성

핵심 포인트

- 체계적 문헌고찰의 목적은 보건정책 의사결정권자가 필요로 하는 양질의 최신 정보를 제공해 주는 것이다.
- 체계적 문헌고찰은 프로토콜을 통해 사전에 정한 연구 질문과 방법을 사용해야하고 신뢰성 있는 연구들의 결과에 근거함으로써 비뚤림을 최소화해야 한다.
- 체계적 문헌고찰은 잠재적 이해상충으로부터 자유로운 해당 분야의 전문가와 방법론 전문가에 의해 수행되어야 한다.
- 중재의 이해당사자(중재에 대한 의사결정을 하거나 중재에 영향을 받는 사람)는 체계적 문헌고찰의 주요한 결정에 참여할 수 있다.

1.1 체계적 문헌고찰을 하는 이유

체계적 문헌고찰은 사람들의 삶에 영향을 미치는 의사결정이 그와 관련된 최신 연구 근거에 바탕을 두고 이루어져야 한다는 요구에서 시작하였다. 그러나 문헌의 양이 계속 증가하고 있기 때문에, 의사결정자가 적절한 의사결정을 내리기 위해 방대한 일차연구들을 평가하는 것은 불가능하다. 체계적 문헌고찰은 일차연구들을 체계적으로 평가함으로써 중재, 진단, 예후 및 기타 여러 분야에서 최신의 연구지식을 제공하는 것을 지향한다. 체계적 문헌고찰의 수행 과정에 있어서 임기응변식 검색, 검색된 연구의 선정·배제 과정 오류, 다양한 비뚤림의 위험과 같은 문제점이 발생할 수 있기 때문에 이를 최소화할 수 있는 방법을 적용하여 연구들을 요약하고 결과를 합성하여 제시해야 한다.

체계적 문헌고찰은 특정 연구 질문에 대답할 수 있도록 사전에 정해진 포함 기준에 맞는 모든 경험적 근거를 모으려고 시도한다. 비뚤림을 최소화하기 위해 명백하고, 체계적인 방법을 사용한다. 여기에는 사전 연구 질문, 명료한 문헌고찰의 범위, 연구의 포함기준, 적절한 연구를 찾기 위한 노력, 비뚤림위험 평가, 공정하고 객관적인 방법으로 결론을 이끌어내기 위한 분석의 방법 등을 망라한다.

체계적 문헌고찰은 중재의 이득과 위해 크기의 정보와 해당 정보의 확실성을 사용자에게 제공하는 것을 목표로 한다. 서로 다른 중재의 비교결과를 제공함으로써 어떤 중재를 선택하는 것이 특정 상황에서 가장 적절한 중재인지를 선택하게 할 수도 있다. 이렇듯이 체계적 문헌고찰의 일차 목표는 건강이나 보건의료 관련 의사결정에 정보를 제공하는 것이다.

체계적 문헌고찰이 필요한 다른 이유로 새로운 연구를 계획할 때 이미 기존에 같은 내용의 연구가 존재해서 계획 중인 연구가 필요 없는 것은 아닌지 살피는 목적도 있다. 따라서 새로운 일차 연구를 착수하기 전에 관련된 체계적 문헌고찰을 예비 검색해서 검토할 필요가 있다. 이러한 예비 검색을 통해 진행 중인 연구들, 현존하는 지식의 격차, 근거가 부족한 부분(예: 현존하는 연구 근거들이 중요한 연구결과물을 간과하는 부분은 없는지) 등을 파악할 수 있게 해준다. 체계적 문헌고찰은 기존 연구들의 제한점을 파악할 수 있게 함으로써 새로운 연구를 할 수 있는 계기를 제공하기도 한다.

체계적 문헌고찰은 해당 주제의 중요한 분야에서 현존하는 정보를 제공하고 앞으로의 연구를 위한 우선순위를 확인하는 기회를 제공하기도 한다. 때로는 정치적 이슈가 되기도 하고 언론의 관심을 받는 경우도 있다. 이러한 영향력 있는 체계적 문헌고찰을 완료하는 것은 종종 까다롭고 시간이 오래 걸린다.

1.2 핵심 질문

연구 질문을 제대로 설정하는 것은 체계적 문헌고찰의 성공적인 수행의 핵심이다. 체계적 문헌고찰 저자들은 핵심 질문이 해당 문헌고찰의 사용자들에게 중요한 것인지를 고려해야 한다.

중재 효과에 대한 질문을 만들 때는 인구집단(population), 중재(intervention), 비교중재(comparison), 중재결과(outcome)로 구성된 PICO방법이 흔히 사용된다. 체계적 문헌고찰의 특성상 중재의 구체적 수행방법, 경제학적 문제들, 형평성, 환자의 경험 같은 추가 질문이 제기될 수도 있다.

1.3 체계적 문헌고찰의 수행자

체계적 문헌고찰은 개별 연구자 보다는 연구팀에 의해 수행되어야 하며, 한 명의 연구자가 체계적 문헌고찰을 출판하는 경우는 흔하지 않다. 팀으로 일한다는 것은 단지 업무를 나눈다는 차원만이 아니라 적격(eligible) 연구의 선정, 자료의 추출, 근거 수준의 평가 등 많은 부분에서 두 명 이상의 연구자가 독립적으로 개입함으로써 비뚤림을 최소화해야 한다는 의미이다. 처음에 참여하는 저자는 이미 경험이 있는 다른 저자에게서 적절한 훈련을 받을 것을 권장한다.

연구팀에는 대상 주제의 전문가를 포함해야 하며, 관련된 모든 관점을 고려해야 하기 때문에 지나치게 협의의 관점을 가진 전문가는 적절하지 않다. 서로 다른 분야의 전문가들이 참여함으로써 단지 하나의 전문성에만 의존할 때 나타날 수 있는 가정이나 용어의 지나친 편협성을 피할 수 있다. 팀 내에는 반드시 통계전문가를 포함한 체계적 문헌고찰 방법론 전문가가 포함되어 있어야 한다.

체계적 문헌고찰 방법론 전문가만으로 중재에 대한 문헌고찰의 수행이 충분히 가능하며 선입견의 위험을 피하기 위해 오히려 해당 분야의 전문가는 배제하는 것이 바람직하다는 주장

도 있지만, 해당 주제 분야와 방법론 전문가가 함께 있을 때 주제 전문가가 중재의 수행, 중재 대상자에 대한 본질적 이해, 여러 중재결과들 사이의 관련성, 의사결정에 영향을 미칠 수 있는 다른 요인들에 대해 중요한 인식을 제공함으로써 기술, 지식, 객관성의 적절한 혼합을 가능하게 한다.

체계적 문헌고찰을 수행할 때는 근거에 대해 독립적으로 평가를 해야 하며 재정적 또는 비재정적인 이해상충을 피하기 위한 주의가 필요하다. 또한, 실재하거나 인지할 수 있는 이해상충이 존재하는지를 고려하는 것이 중요하다. 고용, 특허, 기타 재정적인 지원 등과 관련 있는 연구자는 연구팀에 포함되어서는 안 된다. 이와 같은 문제를 시작 전에 인지할 수 있도록 체계적 문헌고찰에 참여할 계획인 연구자들은 이해상충에 대한 자문을 받아야 한다. 저자들은 연구 시작단계에서 이해상충 선언을 해야 하며 문헌고찰의 각 과정(주제 결정, 프로토콜 개발, 체계적 문헌고찰 수행, 갱신) 또는 자신의 신변에 중요한 변화가 있을 때마다 이를 반복해야 한다.

1.3.1 소비자 및 다른 이해 관계자 참여

의사 결정자와 소비자가 생각하는 우선순위가 연구자와는 다를 수 있기 때문에 저자는 다양한 이해당사자들에게 중요한 질문이 무엇인지를 신중하게 고려하는 것이 중요하다. 대상 주제와 방법론 측면에서 다양한 경험을 가진 사람들이 참여한다면 체계적 문헌고찰은 광범위한 최종사용자들에게 적절히 이용될 수 있을 것이다. 소비자뿐만 아니라 정책 결정자, 연구비 지원자, 의료전문가 같은 더 많은 여러 이해 당사자가 참여할수록 문헌고찰의 적절성은 증대되고 상호 이해가 높아지며 문헌고찰의 수용성은 촉진될 것이며 가용자원의 소모도 줄일 수 있을 것이다.

체계적 문헌고찰 질문과 관련된 모든 잠재적 이해당사자들의 범위를 정하고 누구를 문헌고찰과정에 초대할 것인지를 정하는 것이 첫 번째 단계이다. 이해당사자는 환자, 소비자, 소비자의 대리인, 정책결정자, 공무원, 지침개발자, 전문가 단체, 연구자, 연구비 지원자, 의료인, 기자 등이 포함되며, 연령대의 균형, 신뢰성, 다양성 등도 고려해야 한다. 문헌고찰 저자는 가용자원이 제한적인 나라 또는 지역의 요구도 고려해야 하며 문헌고찰의 범위와 질문에 적절한 참여를 요청해야 한다.

체계적 문헌고찰을 포함해 여러 보건 관련 연구에 소비자가 참여하는 것은 바람직한 행태이며, 여기에서 '소비자'란 환자나 개인적 차원의 의료경험자, 공공서비스의 사용자나 그 일원이라고 할 수 있는 광범위한 범위의 사람들을 말한다. 소비자가 참여하는 체계적 문헌고찰은 그렇지 않은 문헌고찰에 비해 의사 결정자에게보다 더 직접적으로 적용될 수 있다.

1.3.2 소비자 및 기타 이해 관계자와 협력

소비자 및 기타 이해 관계자와 함께 일하는 방법에는 설문 조사, 워크샵, 포커스 그룹 및 자문 그룹 참여 등이 있다. 어떤 방법으로 참여시킬지는 일반적으로 자원 가용성에 따라 결정하면 되지만 문헌고찰 팀은 각 방법의 장단점을 알고 있어야 한다. 저자는 참여할 사람과 방법을 결정해야 하며 참여를 위해 적절한 지원을 제공해야 한다. 지원에는 재정적 보상, 훈련의 제공 등이 있다.

소수의 소비자 또는 이해 당사자가 문헌고찰 팀의 일원이 되어 문헌고찰의 공동 저자가 될 수 있지만, 때로는 모든 사람이 더 넓은 범위의 관점을 제시하고 저자가 되는 역량을 보이거나 이에 관심을 두는 것은 아니라는 것을 인식하는 것이 중요하다. 자문은 소비자 및 관련 이해 관계자, 특히 의견이 다른 주제에 대해 편리한 참여방식을 제공한다. 성공적인 참여를 위한 주요 사항은 다음과 같다.

- 문헌고찰 팀은 주요한 문헌고찰 관련 의사결정을 자문단에게 알려야 한다.
- 적절한 체계적 문헌고찰을 위해 자문은 체계적 문헌고찰 수행 내내 계속되어야 한다.
- 자문단 구성원은 문헌고찰 질문의 폭을 반영해야 하며, 중재의 가치에 대한 결론이 사회의 모든 그룹에 잘 알려 지고 적용될 수 있도록 취약 계층 참여를 고려해야 한다.

1.4 신뢰의 중요성

체계적 문헌고찰은 주어진 문제에 대해서 현재의 지식 상태를 정확하게 표현하는 것을 목표로 한다. 문제에 대한 이해가 증진됨에 따라 문헌고찰을 업데이트 할 수 있다. 그렇지만, 문헌고찰 자체가 출판 당시에 정확해야 한다. 이 정확성이 필수적인 두 가지 이유가 있다. 첫째, 사람들의 건강관련 의사결정이 체계적 문헌고찰 결과를 토대로 하는 경우가 점점 더 늘어나고 있기 때문이다. 현재의 지식은 불완전할 수 있지만 최선의 현재지식이란 관점에서 받아들여질 때 더 나은 의사결정이 가능하다. 둘째, 체계적 문헌고찰은 약물 제품허가, 보험 급여와 같은 규제의 핵심 요소를 형성하고 있다. 여기서, 체계적 문헌고찰은 규제 심사를 위한 감사 가능한 프로세스로 유지되어야 한다. 체계적 문헌고찰은 정확해야 하며 상세한 근거기반 방법을 통해 감사 가능한 정보를 제공할 수 있어야 한다.

1.5 프로토콜 개발

체계적 문헌고찰에 대한 준비는 복잡하며 많은 판단이 필요하다. 문헌고찰 과정에서 잠재적 비뚤림을 최소화하려면 이러한 판단은 가능한 한 문헌고찰에 포함된 연구들의 결과에 의존

하지 않는 방법으로 이루어져야 한다. 문헌고찰 저자의 근거에 대한 사전 지식에 의해 영향을 받을 수 있는 것에는 핵심질문의 정의, 연구 포함기준의 선택, 중재와 비교중재, 중재결과의 사전 정의 등이 있다. 이들을 사전에 결정하고 이를 문서화 하는 것이 중요하다(MECIR C 19-23).

체계적 문헌고찰에서 문헌고찰 질문에 따라 채택할 연구 방법을 결정해야 한다. 질문은 특정 중재와 다른 중재간의 비교 같이 상대적으로 직접적인 것일 수도 있고, 네트워크 메타분석의 일부로 다양한 중재들을 비교하는 계획이 필요할 수도 있으며, 서로 다른 인구 집단, 서로 다른 방법으로 전달된 중재의 다양한 효과를 평가하는 것일 수도 있다.

프로토콜의 핵심 목적은 문헌고찰의 최종 결과에서 비뚤림을 최소화하는 계획을 세우는 것이다. 믿을 수 있는 근거의 합성을 위해서는 계획적이며, 체계적 접근이 필요하다. 체계적 문헌고찰의 타당성을 위협하는 요인들은 문헌고찰이 포함하는 연구들이나 문헌고찰을 수행하는 과정에서 발생한다. 각 연구의 비뚤림은 연구대상자 배정, 눈가림, 자료수집, 자료분석, 연구결과 보고 등에서 나타날 수 있다. 이 같은 문제들을 평가하는 방법이 사전에 프로토콜 상에 결정되어 있어야 한다. 문헌고찰 과정 중에 비뚤림이 없는 연구들을 찾아내는데 실패한다면 비뚤림이 발생할 수 있다. 또한 다루는 중재결과의 속성에 따라 선택 가능한 연구들의 범위가 영향을 받을 수 있다.

체계적 문헌고찰을 위한 프로토콜을 개발하면 비뚤림을 줄이는 것 이상의 이점을 갖는다. 체계적 문헌고찰의 설계에 노력을 더 기울이면 수행 과정이 보다 관리가능해지며, 주요 우선순위를 결정하는 데도 도움이 될 것이다. 질문을 정의하고, 문제를 해결하기 위한 적절한 방법을 사용하게 되면, 분석과 보고에 초점을 맞추고, 문헌고찰이 자금 제공자, 정책 결정자, 의료 전문가 및 소비자를 위한 치료 결정에 가장 많은 정보를 제공할 수 있도록 보장한다. 중재 효과 추정치의 확신도에 영향을 미치는 요인의 체계적 고려를 통해 결과를 해석하는 방법과 함께 연구 전반에 걸친 이질성 조사를 포함한 분석 계획의 세부사항을 프로토콜에 명시해야 한다(MECIR C21-23).

체계적 문헌고찰을 수행하는 중에는 프로토콜을 준수하여야 하지만, 프로토콜의 변경이 때때로 필요하다. 이것은 무작위배정 임상시험의 프로토콜에서도 해당하는 경우인데, 프로토콜은 연구대상자 모집, 자료수집 또는 중재결과 발생과 같은 부분에서 예상치 못한 상황에 적응하기 위해 변경될 수 있다. 사전에 결정된 프로토콜을 준수하기 위해 모든 노력을 기울여야 하지만, 이것이 항상 가능하거나 적절한 것은 아니다. 다만, 임상시험이든 체계적 문헌고찰이든 연구결과에 어떤 영향을 미치는가에 근거해 프로토콜의 변경이 이루어져서는 안 된다는 것이 중요하다. 체계적 문헌고찰에서 선택된 연구를 배제하거나 통계 분석을 변경하는 등 연구결과에 미치는 영향이 알려진 경우 행해진 사후 의사결정은 비뚤림에 매우 취약하므로 이에 대한 합리적인 근거가 없는 한 피해야 한다.

1.6 데이터 관리 및 품질 보증

체계적 문헌고찰은 재현 가능해야 하며, 연구의 포함 여부 결정, 자료수집, 자료변환 또는 자료수정의 기록을 유지하는 것이 안전하고 감사가능하게 하는 데 도움이 될 것이다. 자료가 부주의하게 손실되지 않도록 하고 오류의 식별 및 수정을 용이하게 하여야 한다. 또한 향후의 업데이트 및 유지보수를 지원하기 위해서는 좋은 자료 관리 프로세스가 필수적이다. 문헌고찰 결정 과정의 투명한 보고를 통해 독자는 스스로 문헌고찰의 신뢰성을 평가할 수 있다.

Covidence 및 EPPI-Reviewer와 같은 관리 소프트웨어를 사용하여 문헌고찰의 결과로 내린 결정과정을 일관되고 표준화된 기록의 형태로 유지할 수 있다. 이러한 도구는 문헌고찰 팀 구성원이 전 세계에서 원격으로 액세스할 수 있는 자료의 중앙 저장소를 제공한다. 이 프로그램은 자료의 선택, 비뚤림평가, 자료추출에 대한 연구의 독립적 평가를 기록하여 필요한 경우 프로세스 후반에 점검할 수 있도록 한다. 연구에 따르면 경험 많은 문헌고찰 저자들조차 실수를 하고 비뚤림위험성 평가가 서로 다를 수 있기 때문에 저자들이 시간을 더 투자해야 한다는 문제는 있지만 품질을 유지하는 것이 특히 중요하다. 문헌고찰에서 보다 정교한 정보 기술 도구들이 사용되기 시작하면서, 포함 여부에 대한 결정을 포함한 모든 자료가 개별차원의 문헌고찰의 범위를 넘어서는 가치를 갖는 것이 점점 명백해지고 있다. 예를 들어, 기계 학습을 위해 원래 문헌고찰의 자료를 사용할 수 있는 경우 반자동 또는 자동화를 통해 추후 문헌고찰의 업데이트를 보다 효율적으로 수행할 수 있다.

1.7 연구 기획 관련 MECIR

C19: 프로토콜 단계에서 문헌고찰 방법을 기획함(필수)	
연구들을 파악할 방법을 미리 계획한다. 적절한 출판시기 및 자료원을 충족하면서 언어나 출판형태에는 제한을 두지 않은 채 선택기준을 충족하는 연구를 가능한 한 많이 찾아낼 수 있는 검색을 디자인한다.	검색은 문헌고찰의 선택기준에 직접 근거하여 이루어져야 한다. 프로토콜을 기획할 때에는 선택기준을 충족하는 모든 종류의 연구를 대상으로 하는 것이 중요하다. 만일 출판형태나 언어에 따라 검색에 제한을 하는 경우 출판 비뚤림이나 언어 비뚤림이 생겨날 가능성이 있다. 영어 데이터베이스에서 언어제한을 제거한다고 비영어 잡지검색과 비영어 데이터베이스 검색을 대체할 수는 없다.
C20: 포함된 연구에서 비뚤림의 위험성 평가 계획(필수)	
포함된 연구들의 비뚤림위험평가에 사용될 사용도구 및 이를 어떻게 사용할 것인지, 낮은/높은/불명확한 비뚤림위험 같이 각 연구를 판단할 기준을 미리 계획한다.	비뚤림위험 평가의 기준을 미리 정하는 것은 중요하다. 왜냐하면 문헌고찰에서 나온 결과를 분석하거나 해석하는 것은 이 기준에 의한 판단에 영향을 받기 때문이다. 무작위시험에서는 코크란의 비뚤림위험 평가를 이용하는 것이 필수이다. 이 도구는 비뚤림위험 평가를 할 수 있는 충분하면서도 쉬운 도구이다.
C21: 결과합성의 기획(필수)	
포함된 연구의 결과를 종합하는 데 사용할 방법을 미리 계획한다. 여기에는 정량적 합성이 계획되어 있는지 여부, 이질성을 평가할 방법, 효과 측정 선택(예: 이분형 변수를 대상으로 구하는 오즈비, 위험비, 위험차 및 기타 다른 방법들), 메타분석 방법(예: 역분산법 또는 맨텔-핸젤 방법, 고정 효과 또는 무작위 효과 모형).	문헌고찰 결과의 분석이나 해석은 이 과정에서 이루어진 판단에 영향을 받을 수 있으므로 합성 방법, 특히 통계 방법을 미리 정의하는 것은 중요하다.
C22: 하위군 분석의 기획(필수)	
프로토콜 단계에서 잠재적 효과 변경요인(예: 하위군 분석을 위한)를 미리 정의하고, 하위군의 개수를 제한하고, 각각에 대한 근거를 제공한다.	사전에 정의함으로써 다수의 방향성 없는 하위군 분석이 이루어져 이질성에 대해 의미 없는 설명을 하는 위험을 감소시킨다.
C23: GRADE 평가를 기획하고 결과요약표를 기획(필수)	

근거 수준을 평가하고 문헌고찰 결과를 요약하는 데 사용할 방법을 미리 계획한다.	고찰에서 가장 중요한 중재결과에 대한 근거 수준을 평가하는 방법을 사전에 정해야 한다. '결과요약' 표에서 가장 중요한 특징은 문헌고찰에서 중재결과의 선택적 제시를 막기 위해 중재결과를 미리 정한다는 것이다. 결과 요약표에는 의사결정에 필수적인 필수 중재결과(일반적으로 최대 7개)를 포함하며, 대리 또는 중간 중재결과는 포함하지 않는다. 중재결과를 선택할 때 예상되거나 관찰된 효과의 크기에 기초해서 하면 안 된다.

1.8 참고문헌과 읽을거리

- Chalmers I, Bracken MB, Djulbegovic B, Garattini S, Grant J, Gulmezoglu AM, Howells DW, Ioannidis JP, Oliver S. How to increase value and reduce waste when research priorities are set. Lancet 2014; 383: 156–165.
- Chandler J, Hopewell S. Cochrane methods – twenty years experience in developing systematic review methods. Systematic Reviews 2013; 2: 76.
- Lasserson TJ, Thomas J, Higgins JPT. Chapter 1: Starting a review. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA(editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.0(updated July 2019). Cochrane, 2019. Available from www.training.cochrane.org/handbook.
- Moher D, Liberati A, Tetzlaff J, Altman D, PRISMA Group. Preferred reporting items

제2장 주제범위 결정

핵심 포인트

- 체계적 문헌고찰은 대답가능한 질문을 다루어야 하며 지식의 중요한 간극을 해소해야 한다.
- 바람직한 질문을 만들기 위해서는 시간과 전문 지식, 문헌고찰 사용자의 참여가 필요하다.
- 질문은 광범위할 수도 좁을 수도 있으며 각각의 장단점이 있다.
- 로직모델(logic model)은 중재, 특히 복잡한 중재가 어떻게 작동하는지를 문서화하는 방법이며, 문헌고찰의 질문 및 광범위한 범위를 구체화하는 데 사용될 수 있다.
- 관련 이해당사자가 우선순위 설정, 형평성 검토 등에 참여한다면 문헌고찰이 범위와 초점에 따른 올바른 질문을 다루도록 도움을 줄 수 있다.

2.1 잘 만들어진 질문의 필요성

다른 연구와 마찬가지로 체계적 문헌고찰을 준비하는 데 있어서 우선적이고 중요한 것은 연구의 초점을 결정하는 것이다. 이것은 체계적 문헌고찰이 답하고자 하는 질문의 틀을 명확하게 함으로써 가능해 진다. 체계적 문헌고찰의 초점은 건강이나 건강관련 의사결정을 내리는 사람들에게 중요한 질문이어야 한다는 것이다. 일반적으로 이러한 의사결정을 위해서는 중재의 이득과 위해를 모두 고려해야 한다(MECIR C1, C3). 좋은 문헌고찰 질문을 만들기 위해서는 시간이 걸리는 경우가 많으며, 주제 영역뿐만 아니라 광범위한 이해 관계자 그룹의 참여가 필요하다.

잘 구성된 질문을 통해 적격 기준 결정, 연구 검색, 포함된 연구에서 자료추출, 합성, 및 중재결과 제시를 포함한 문헌고찰과정의 많은 과정들이 잘 진행될 수 있다. 체계적 문헌고찰에서 질문은 광범위하게는 문헌고찰의 '목표'로 명시되며, 연구의 운영 측면에서는 '이 문헌고찰에 포함할 연구인지를 고려하는 기준'으로 사용할 수 있다. 이 절에서는 문헌고찰의 초점을 결정하는 문제를 다룬다. 문헌고찰 질문은 또한 현재의 문헌고찰이 독자들이 직면한 문제와 직접적으로 관련될 가능성이 있는지에 대해서도 유용한 정보를 제공한다.

FINER 기준은 연구 질문을 개발할 때 다루어야 할 이슈를 요약하기 위해 제안되었다. 이 기준에 따르면 질문은 실현가능하고(feasible), 흥미로워야 하며(interesting), 참신하고(novel), 윤리적이며(ethical) 및 관련성이 있어야(relevant) 한다. 이러한 모든 기준은 문헌고찰을 시작할 때 그리고 질문을 만들 때 모두 염두에 두고 있어야 한다.

실현가능한 문헌고찰이란 이용가능 한 근거를 사용하여 다룰 수 있는 질문을 하는 것이다. 실현가능성측면에서 질문에 대한 검색의 결과가 관리가 불가능한 분량이 되지 않게 하는 것이 중요하다. 사전 범위 조정 작업을 통해 문헌고찰을 위한 합리적인 경계를 정의해야 한다. 마찬가지로 근거에 기반 해서 간극(gap)을 식별하는 것이 유용할 수 있지만, 체계적 문헌고찰 저자와 이해당사자는 기존 근거로는 답변할 수 없는 질문을 하는 것인지를 인식해야 한다.

문헌고찰에 참여하는 저자는 해당 분야에 흥미를 가지고 있는지 물어보아야 한다. 문헌고찰은 의미 있는 과업이고 문헌고찰의 저자는 해당 작업을 끝까지 함께 할 수 있는 충분한 책임감이 필요하기 때문에 저자가 관심을 갖는 주제에 대해 문헌고찰을 수행하는 것이 중요하다.

참신한 문헌고찰은 지식의 실제적 간극을 해소할 것이므로, 관련되거나 중복된 문헌고찰이 존재하는지를 알아야 한다. 이것은 노력의 중복을 줄이고 또한 저자들이 문헌고찰이 기여할 더 넓은 연구 맥락을 이해하도록 보장한다. 문헌고찰을 시작하기 전에 발표된 연구 문헌에 존재하는 기존의 체계적 문헌고찰이 있는지를 PROSPERO 등록부에서 지속적으로 확인해야 한다.

체계적 문헌고찰처럼 까다로운 활동을 수행하는 데 수반되는 기회비용을 고려해보았을 때 저자는 자신의 문헌고찰 작업이 (1) 초점과 문제를 정의할 때 이해당사자가 참여하고 있는지, (2) 문헌고찰의 연구결과를 바탕으로 의사결정을 할 때 용이하게 할 수 있도록 관련성이 있는지 확인해야 한다.

윤리적 문제는 우선순위와 관련된 질문에서 주로 고려해야 하지만, 윤리적 측면에서도 기회비용에 대한 고려는 이루어질 수 있다. 연구 질문은 종종 가치중립적이지 않으며, 주어진 문제에 접근하는 방법은 예를 들어 의도적이든 아니든 결과적으로 건강 불평등의 확대와 같은 문제를 초래할 수 있는 정치적 의미를 가질 수 있기 때문이다.

2.2 중재연구 문헌고찰의 목표

체계적 문헌고찰은 일차연구로 대답할 수 있는 모든 질문을 다룰 수 있다. 중재의 효과를 검토하는 체계적 문헌고찰은 상당히 다양한 목표를 가질 수 있다. 어떤 사람들은 특정 중재와 이의 대안 중재에 대한 근거에 특히 초점을 맞추지만, 다른 사람들은 다양한 중재들의 비교에 관심을 가질 수도 있다. 여러 가지 중재를 문헌고찰하고 어떤 것이 가장 효과적인지 식별하는 것을 목표로 하는 문헌고찰은 단일 중재를 문헌고찰 하는 연구자들의 작업에 비해 더 광범위하고 어려울 수 있다. 하지만 이는 최종 사용자에게 가장 유용할 수 있으며, 의사결정에는 여러 가지 중재 옵션 중에서 선택하는 것이 포함될 수 있다.

특정 중재에 기인한 이득과 위해의 균형을 살펴보는 것뿐만 아니라, 중재 효과의 크기와 인구적인 측면, 중재 자체, 어떻게 중재결과를 측정했는가와 같은 다른 특성, 그리고 일차 연구의 방법론사이의 관계를 조사하는 것이 목표가 될 수 있다. 그러한 접근방식은 특정 다요소 중재 중 어떤 구성요소가 더 중요하거나 덜 중요한지 그리고 언제 중요한지를 조사하는데 사용될 수 있다. 유용한 중재가 어떻게 그 효과를 달성하는지를 항상 알 필요는 없지만, 많은 문헌

고찰은 이것을 문헌고찰의 명시적 목적으로 하거나 문헌고찰의 범위를 설명할 때 중재의 작용 기전을 명확히 하는 것을 목표로 할 것이다. 중재가 어떻게 작용하는지를 이해하는 것은 의사결정자들로 하여금 문헌고찰이 자신들의 상황에 대한 적용가능한지를 평가하는 데 중요한 도움이 될 수 있다. 또한 많은 의사결정이 자원의 이용 가능성에 의해 적어도 부분적인 제약이 존재하므로 문헌고찰 저자들은 종종 중재의 경제적 맥락을 고려할 필요가 있다.

2.3 문헌고찰 질문의 범위를 정의하기

치료적 중재를 비교하는 무작위배정 임상시험 같은 연구는 서로 다른 중재에 따른 중재결과를 비교한다. 통계적 합성(예: 메타분석)은 새로운 중재 대 관행적 기존 중재 또는 경쟁적인 두 중재의 비교에 초점을 맞춘다. 여기서 우리는 '실험 중재' 대 '대조 중재'란 용어를 사용한다. 이는 중재 중 하나를 실험으로 식별할 필요성에 따른 것이고, 양군 모두 통제되고, 일대일로 비교되기 때문에 단지 편의상 구분일 뿐이다. 두 군간의 중재결과 비교치를 '중재 효과'라고 한다.

문헌고찰의 목적을 기술하려면 문헌고찰의 일차 목적을 정밀하게 정의하면서 시작해야 하며, 이상적으로는 한 문장으로 해야 한다(MECIR C2). 가능한 경우 스타일은 '[특정 사람, 질병 또는 문제, 세팅]에서 [건강 문제]에 대한 [중재 또는 비교]의 효과를 평가하는 형식이어야 한다. 이어서 다른 연구대상자, 다른 중재 및 비교, 또는 다른 중재결과와 관련된 하나 이상의 이차적 목표가 뒤따를 수 있다. 문헌고찰 질문의 세부 사양은 인구, 중재, 비교 및 중재결과의 약자인 'PICO'로 표현할 수 있다. 각 PICO 구성요소를 지정할 때 각각의 요소를 동일하게 강조할 필요는 없다. 예를 들어, 유방암의 특정 단계에 대한 경쟁적 중재에 초점을 맞춘다면 질병의 단계와 중증도는 매우 정밀하게 정의할 수 있으며, 반대로 어떤 병기의 유방암에서라도 사용되는 특정 치료약물이라면 치료 제형이 매우 정밀하게 정의되는데 초점을 맞출 수 있다.

PICO 구성은 문헌고찰의 세 가지 다른 단계에서 이루어지며 이들은 구분이 필요하다.

- **문헌고찰 PICO**(프로토콜 단계에서 계획됨)는 연구의 적격성을 기초로 하는 PICO이다(어떤 연구를 포함하고 배제할 것인가).
- **각각의 합성에 대한 PICO**(프로토콜 단계에서 계획됨)는 각 양적 합성이 어떻게 구성될지를 결정하고 계획된 비교(중재 및 대조군, 중재결과 및 하위군 그룹화 포함)를 명시하는 것을 목표로 하는 질문을 정의한다.
- **포함된 연구의 PICO**(문헌고찰 단계에서 결정됨)는 포함된 연구에서 실제로 조사된 것이다.

문헌고찰 PICO가 연구의 목적을 명확히 할 수 있는 지점에 도달하려면 문헌고찰 저자와 사용자 간에 세부적인 논의가 필요하다. 문헌고찰의 범위와 질문 개발에 관여하는 사람들은 문헌고찰에서 다루게 될 실질적인 문제와 양적 합성이 이루어질 연구 분야를 잘 알고 있어야 한

다. 그 질문들을 개발하는 것은 연구 과정의 중요한 부분이다. 이와 관련해서 몇 가지 방법론적인 이슈가 있는데 어떤 질문이 대답이 필요한 가장 중요한 것인지를 결정하는 방법, 질문 공식에 이해당사자를 참여시키는 방법, 문헌고찰이 진행됨에 따라 초점의 변화를 어떻게 다루어야 하는지, 문헌고찰을 얼마나 광범위하게 또는 좁게 해야 하는지에 대한 고려사항 등이 그런 것들이다.

2.3.1 광범위 문헌고찰 대 좁은 문헌고찰

문헌고찰에서 다루는 질문은 범위가 넓거나 좁을 수 있다. 예를 들어, "항혈소판제가 모든 혈전증을 방지하는데 효과적인지"와 같은 광범위한 질문을 다룰 수 있다. 또는 "아스피린과 같은 특정 혈소판 억제제가 이전의 뇌졸중 병력이 있는 노인들의 특정 혈전증 사건 예들 들어 뇌졸중의 위험을 줄이는 데 효과적인지"에 대해 문헌고찰 할 수 있다. 다양한 치료 옵션 중에서 어떤 중재가 가장 효과적인지 또는 중재수행과 연구대상자 특성에 따라 어떻게 중재결과가 다른지 조사하기 위해, 문헌고찰은 점점 더 광범위해지고 있다.

여러 개의 문헌고찰을 요약하는 문헌고찰 오버뷰(overview of reviews)는 동일한 상태의 연구대상자에게 적용한 서로 다른 중재의 효과에 대한 복수의 문헌고찰 또는 다양한 상태의 연구대상자에게 적용한 동일한 중재에 대한 복수의 문헌고찰을 요약할 수 있기 때문에 광범위한 근거 기반 요약의 필요성을 충족하는 한 가지 방법이 될 수 있다. 이와 함께 비교적 좁은 범위의 문헌고찰을 연속적으로 진행하는 것도 바람직하다고 생각할 수 있다. 대안으로서, 동일한 문헌고찰 내에서 서로 다른 치료 옵션의 비교를 결합하는 것이 특히 네트워크 메타분석 방법의 성장을 고려할 때 더 유용할 수도 있다. 문헌고찰이 가장 적절한 접근방식인지 여부를 결정할 때 문헌고찰 저자는 질문의 폭과 가용 자원을 고려해야 한다. 일부 질문은 모든 관련 일차 연구를 대상으로 문헌고찰을 실행가능하기엔 너무 광범위하며, 만약 어떤 분야가 충분한 양질의 문헌고찰을 가지고 있다면, 일차 연구를 대상으로 문헌고찰을 다시 시행하는 것은 합리적인 자원의 사용이 아닐 수 있다.

광범위 문헌고찰과 좁은 문헌고찰의 장단점은 표 2.1에 요약되어 있다. 연구대상자의 범위 측면에서 넓은 범위를 갖는 것은 일반화 가능성을 높일 수 있는 잠재성을 가지고 있지만, 발견이 궁극적으로 더 넓은(또는 다른) 인구집단에 적용될 수 있는 정도는 실제로 연구에 모집된 연구대상자에 따라 달라질 것이다. 마찬가지로, 이질성은 연구 간 효과의 동질성을 기대한다면 단점이 될 수 있지만, 문헌고찰의 질문이 차등적인 효과에 대한 이해를 지향한다면 반대로 장점이 될 수 있다.

문헌고찰의 폭과 그 문헌고찰 내부 질문의 정밀성은 구분해야 할 필요가 있다. 광범위한 문헌고찰이라 하더라도 상당히 좁은 질문을 가질 수 있기 때문이다. 예를 들어, 혈전증을 대상으로 하는 항혈소판제제 투여 효과라는 광범위한 체계적 문헌고찰에 이용 가능한 모든 치료법을 비교할 수도 있다. 그러나 모든 연구를 하나의 비교로 결합하는 것보다, 특정한 치료법을 다른

특정 치료법과 비교하여 더 좁고 더 동질적인 비교로 세분화할 수도 있다. 이는 앞에서 설명한 세 가지 PICO 수준과 관련이 있다. 문헌고찰 PICO는 문헌고찰의 넓은 범위를 정의하고, 비교 PICO는 서로 비교할 특정 치료법을 정의한다. PICO의 사용에 대한 상세한 설명은 3장에서 기술한다.

실제로 체계적 문헌고찰은 넓은 범위에서 시작할 수 있으며, 근거가 축적되고 원래의 문헌고찰이 감당할 수 없게 너무 커지면 더 좁은 문헌고찰로 나눌 수 있다. 이는 실제적이고 관리적인 이유 때문에 일어날 수 있는데 이렇게 하면 문헌고찰의 갱신이 용이해지고 독자들이 근거에 기반해서 어떤 부분이 변경되고 있는지를 보다 쉽게 확인할 수 있기 때문이다. 개별 문헌고찰 저자는 광범위한 문헌고찰을 일련의 좁은 문헌고찰로 나누는 것이 적절한지 여부를 결정하고 이를 달성하기 위한 적절한 방법을 구현해야 한다. 광범위한 문헌고찰을 보다 좁은 일련의 문헌고찰로 분할하는 등 중대한 변경을 수행하려면 각 문헌고찰별 대상 연구의 적격 기준을 문서화하는 새로운 프로토콜을 작성해야 한다.

궁극적으로 문헌고찰 범위의 선택은 질문의 목적 적합성과 잠재적 영향, 이론적, 생물학적 및 역학적 정보 지원, 질문에 대한 답변의 잠재적 일반성과 타당도, 가용 자원을 포함한 여러 요인에 따라 달라진다. 저자는 문헌고찰을 위한 최적의 범위를 결정할 때 문헌고찰 사용자의 요구와 문헌고찰이 사용될 것으로 기대하는 맥락을 신중하게 고려해야 한다.

표 2-1. 광범위 문헌고찰과 좁은 문헌고찰의 장단점

구 분		광범위 문헌고찰	좁은 문헌고찰
연구대상자 선택 예: 어깨건염에 스테로이드 주사(좁음) vs 모든 건염에 스테로이드 주사(넓음)	장점	• 근거의 포괄적인 요약 • 다양한 유형의 연구대상자로 인한 연구결과의 일관성(일반화 가능성)을 탐색할 수 있는 기회	• 고찰 팀의 관리용이성 • 읽기 쉽다
	단점	• 검색, 자료수집, 분석 및 작성에 더 많은 자원을 필요로 함 • 큰 규모의 문헌고찰이지만 이런 다양한 연구대상자를 포함하는 명확한 근거(예: 조사 결과의 일관성)가 결여된 경우, 해석은 독자에게 어려울 수 있음	• 관련 근거가 드묾 • 중재가 다른 상황이나 인구집단에서 다르게 작용하는지 여부를 탐색할 수 없음(예: 불평등을 초래할 수 있는 차등 효과를 탐색할 수 없음) • 복수의 문헌고찰에 접근해야 하는 경우(예: 관심 인구집단에 대한 근거가 드문 경우) 의사결정자의 부담 증가 • 문헌고찰 저자가 원하는 연구결과를 도출하기 위해 범위를 임의로 선택할 가능성.

구 분		광범위 문헌고찰	좁은 문헌고찰
중재 방법 예: 우울증을 위한 감독하의 달리기(좁음) vs 우울증을 위한 모든 종류의 운동(넓음)	장점	• 근거의 포괄적인 요약 • 다양한 중재 수행을 아우르는 연구결과의 일관성을 탐색할 수 있는 기회	• 고찰 팀의 관리용이성 • 읽기 용이
	단점	• 검색, 자료수집, 분석 및 작성에 더 많은 자원을 필요로 함 • 큰 규모의 문헌고찰이지만 이런 다양한 중재를 포함하는 명확한 근거(예: 조사 결과의 일관성)가 결여된 경우, 해석은 독자에게 어려울 수 있음	• 관련 근거가 드묾 • 중재를 다르게 하는 것이 중재의 효과를 변화시키는지의 여부를 탐구할 수 없음 • 복수의 문헌고찰에 접근해야 하는 경우(예: 특정 중재 방법에 대한 근거가 드문 경우) 의사결정자의 부담 증가 • 문헌고찰 저자가 원하는 연구결과를 도출하기 위해 범위를 임의로 선택할 가능성
중재와 대조 중재의 선택 예: 오줌싸개를 방지하기 위한 옥시부티닌과 데스모프레신의 비교(좁음) vs 오줌싸개를 방지하기 위한 중재의 비교(넓음)	장점	• 근거의 포괄적인 요약 • 다양한 중재 옵션의 효과를 비교할 수 있는 기회	• 고찰 팀의 관리용이성 • 목표의 단순성 및 읽기 용이
	단점	• 검색, 자료수집, 분석 및 작성에 더 많은 자원을 필요로 함 • 다루기 번거로우며 문헌고찰 오버뷰로 발표하는 것이 더 적절할 수 있음	• 복수의 문헌고찰에 접근해야 할 수 있으므로 의사결정자의 부담 증가

2.3.2 '광범위' 대 '좁히기'

문헌고찰 PICO로 문헌고찰의 범위를 정할 때는 연구간 이질성에 대한 우려와 비교 PICO에 의해 결정된 다양한 연구의 결과를 결합하는 정당성에 대한 우려를 혼동하지 않는 것이 중요하다.

광범위한 리뷰는 '사과와 오렌지를 섞는 것'이라는 비판을 받아왔다. 하지만 두 가지 개념('광범위 문헌고찰'과 '사과와 오렌지를 섞음')은 서로 다른 문제다. '사과와 오렌지' 비판은 단일 비교 내에서 부적절한 연구의 혼합을 말하며, 상당히 다른 중재, 다른 인구집단, 다른 환경에서의 중재결과들을 무비판적으로 통합하는 것은 의심스러울 수 있다.

불행하게도, 연구가 서로 결합될 수 있을 만큼 유사한지에 대한 결정은 항상 간단하지 않으며, 때로 '광범위'(동일한 분석으로 연구를 결합하는 경우) 또는 '좁히기' 중 하나로 양분되지만 사실은 연속성의 차원에서 보는 것이 바람직하다.

비교 PICO를 정의하고 '광범위' 또는 '좁히기' 중 어느 것이 적절한지를 고려할 때 명심해야 할 다양한 장점과 단점이 있다. '광범위'는 이질성을 설명할 수 있는 요인의 조사를 가능하게 한다. 이러한 조사의 결과는 중재가 다른 인구에서 다르게 작용하는지 여부에 대한 중요한 단서를 제공할 수 있다. 궁극적으로 이러한 유형의 지식은 임상적 의사결정에 유용하다. 그러나 '광범위'는 이질성을 도입할 가능성이 높으며, 이는 항상 우선순위로 설명되지 않으며, 이는 임상적으로 해석하기 어려운 결합 효과를 초래할 수 있다. 비교를 분할하면 그 것 자체로는 유용한 양적 합성을 산출하기 위한 연구가 너무 적다는 위험이 수반된다. 필연적으로 메타분석을 수행하려면 PICO 요소 전반을 아우르는 어느 정도의 수집이 필요하다.

2.4 문헌고찰이 올바른 질문을 하고 있는지를 보장하기

체계적 문헌고찰은 보건의료 관련 의사결정에 사용되기 때문에 문헌고찰의 핵심질문이 의사결정에도 의미를 가지는지 확인해야 한다. 아래의 두 가지 접근방법에 대해 논의한다.

- 기존의 연구 우선순위 설정의 결과를 사용하여 문헌고찰의 질문을 정의한다.
- 기존의 연구 우선순위 설정이 없거나 그 외에 이해관계와 관련된 경우, 이해당사자와 협력하여 문헌고찰 질문을 정의하고 정책과 실천에 대한 관련성을 확립한다.

2.4.1 우선순위 설정을 이용한 문헌고찰 질문의 정의

연구의 우선순위 설정은 연구로 해결할 가장 가치가 있는 불확실성이 무엇인지를 결정하기 위한 일련의 활동이다. 문헌고찰의 범위를 정의하기 위해 연구의 우선순위 설정을 사용하는 것은 문헌고찰이 이해당사자의 목적에 적합하게 하는데 도움이 된다.

연구의 우선순위 설정은 항상 특정 맥락, 환경 및 특정 원칙, 가치 및 선호(구체적으로 표현되어야 함)를 가진 인구집단에서 수행된다. '연구 질문'의 범위와 목적에 대한 이해당사자의 해석이 다를 수 있으며, 이로 인해 해석하기 어려울 수 있는 우선순위가 생길 수 있다. 연구자나 문헌고찰 팀은 연구의 우선순위를 대답 가능한 PICO 연구 질문 형식으로 변환할 필요가 있다고 생각할 수 있으며, 이해당사자 그룹과 함께 질문을 다시 점검하여 그들의 의도를 정확히 반영했는지 여부를 판단하는 것이 유용할 수 있다.

우선순위에 대한 서로 다른 이해당사자 그룹의 견해가 차이 나는 이유가 문헌고찰의 범위를 정의하는 데 도움이 될 수 있다. 이는 특정 인구집단이나 환경을 배제하거나, 하위군 분석을 포함하고 수행하는 것에 대한 의사결정에 특히 중요하다. 실현 가능하다면 체계적 문헌고찰은 의사결정자, 환자/공공 및 실무자와 같은 주요 이해당사자의 우선순위에 기초해야 한다.

2.4.2 문헌고찰 질문 정의에 이해당사자의 참여

관련 연구 우선순위 설정에 대한 자료가 없거나, 매우 구체적인 목적을 위해 체계적 문헌고찰을 실시하고 있는 경우(예를 들어, 진료지침개발을 위해 위탁받은 경우), 연구자는 관련 이해당사자와 협력하여 문헌고찰 질문을 정의해야 한다. 이런 방법은 국가 및 지역 간의 차이가 존재하는 상태에서 보건 시스템과 정책의 효과성을 연구하는 경우에 매우 유용하다. 이러한 차이의 중요성은 이해당사자와의 논의를 통해서만 명백해질 수 있다.

문헌고찰에 대한 이해당사자에는 소비자 또는 환자, 보호자, 서로 다른 직역의 보건 전문가, 정책 의사결정자 및 기타 직종이 포함될 수 있다. 특정 질문에 중요한 이해당사자를 파악하는 것은 핵심질문이 누구에게 영향을 미칠 가능성이 있는지, 그리고 그것이 효과적인 것으로 판명될 경우 중재를 실행(또는 그렇지 않을 경우 중단)할 것으로 예상되는 사람에 따라 달라진다.

이해당사자 참여는 질문의 정의에서 연구결과의 배포에 이르기까지 체계적 문헌고찰의 전체과정 내내 이루어져야 한다. 이해당사자를 참여시키면 적절성을 높이고, 서로 배우는 것이 많아지고, 참여도를 향상시키고, 연구자원의 낭비를 줄인다. 그러나 그러한 관여 자체는 도전적일 수 있기 때문에 문헌고찰 질문을 정의하는 일회성으로만 참여할 수도 있다. 여러 이해당사자가 참여하여 개념화하고 다듬은 문헌고찰 질문은 체계적 문헌고찰에서 다루어야 할 복잡성의 상당 부분을 포착할 수 있다.

2.4.3 문헌고찰 질문 정의 시 형평성 관련 문제 고려

체계적 문헌고찰 과정은 건강의 불평등을 증가시키거나 감소시키는 정치적 활동으로 간주될 수 있다. 의사결정자들은 서비스 제공을 계획할 때 이러한 지식을 고려할 수 있다. 따라서 저자는 사회적 약자에 해당 중재가 어떠한 영향을 미치는지 그러한 중재의 이행 여부나 방법에 따라 건강상의 사회 경제적 불평등이 영향을 받을 수 있는지 여부를 고려해야 한다.

건강 형평성은 건강에서 피할 수 있는 불공평한 차이가 없는 것을 뜻한다. 건강 불평등은 주거, 인종, 민족, 문화, 언어, 직업, 성별, 종교, 교육, 사회경제적 지위, 사회적 자본 및 성적 성향, 연령 및 장애와 같은 기타 특징에 걸쳐 발생할 수 있다. 문헌고찰 질문이 개발될 때에는 건강 형평성에 관련된 문제를 고려해야 한다(MECIR C4)

2.5 문헌고찰의 구조화를 위한 방법 및 도구

문헌고찰의 검토범위를 주의 깊게 개발하는 것이 중요하다. 문헌고찰이 기존 지식에 기여하는 내용에 대한 명확한 이해가 없다면 개념적 모순의 위험에 처할 수 있다. 이는 중재가 특정 중재결과를 내기 위해 작용하는 맥락과 어떻게 상호작용하는지에 대한 중요한 요소를 잘못 파악하는 경우에 발생할 수 있다.

문헌고찰이 개념적으로 일관성을 유지하고 적절한 과정을 거치기 위해 처음부터 핵심 개념에 대한 정의를 잘 만들고, 문헌고찰이 의존하는 이전의 가정이 명확한지를 확인할 필요가 있다. 이러한 이전의 가정에는 문헌고찰 범위에서 특정 인구집단을 고려해야 하는 이유, 중재의 구체적인 방법, 특정 중재결과를 선택하는 이유가 포함된다. 전체적으로 고려할 때, 초기에 시행하는 이러한 개념과 정의에 대한 작업을 통해 개념적 틀이 완성된다.

이 절에서는 문헌고찰의 범위와 주요 개념 간의 관계를 정의하는 데 도움이 될 수 있는 도구를 고려한다. 특히, 중재가 어떻게 중재결과를 초래하는지를 명시한다. 어떤 상황에서는 실행 중인 중재와 관찰되는 중재결과 사이에 긴 일련의 사건들이 발생할 것으로 예상된다. 예를 들어, 학교에서 천식교육이 어린이들의 건강과 복지에 미치는 영향을 조사하는 체계적 문헌고찰이라면 핵심 중재요소들 간의 상호 작용과 서로 다른 학교 환경에 대한 도입, 각 어린이 수준에서의 서로 다른 효과변경요인(effect modifier)들, 그리고 그 중재가 어떻게 어린이 및 가족의 지식에 영향을 미치는지, 어린이의 자기효능과 치료제 순응도, 천식의 중증도, 활동 제한일수, 학교 출석에 미치는 효과, 교육 성취에 따른 원위적(distal) 중재결과와 어린이 보건 및 웰빙 지표 등이 있다.

저자가 문헌고찰의 질문을 정의하고 문헌고찰을 계획할 때 제기된 문제를 고려하는 데 몇 가지 도구가 도움이 될 수 있다. 이러한 도구는 적격 기준을 개발하고 포함된 연구를 분류할 때 유용하다. 여기에는 다음이 포함된다.

1. 관련 중재, 중재결과 또는 인구집단을 분류(또는 그룹화)하는 데 사용할 수 있는 계층 구조적 분류법(taxonimies).
2. 중재 특성에 대한 설명을 검토하고 구조화하기 위한 일반적인 틀(예: 중재의 설명을 위한 TIDieR(Hoffmann 등 2014), 체계적 문헌고찰에서 복잡성의 여러 측면을 설명하는 iCAT_SR(Lewin 등 2017)).
3. 특정 건강 상황에 대해 측정해야 하는 합의된 중재결과를 식별하고 정의하기 위한 핵심 중재결과 세트.

고찰의 특정 측면에 초점을 맞추는 이러한 도구와는 달리, 로직모델은 문헌고찰 수준에서 양적 합성을 계획하고 지도하기 위한 틀을 제공한다.

2.5.1 로직모델(Logic model)

로직모델(개념적 틀 또는 변화 이론이라고도 함)은 중재가 어떻게 작용하는지에 대한 이론을 그래픽으로 표현한 것이다. 로직모델은 중재의 구성요소, 메커니즘(행동의 경로), 산출물(outputs) 및 중재결과를 일련의 사건(선형적일 필요는 없지만)으로 묘사한다. 로직모델은 주로 중재의 복잡성을 다루는 체계적 문헌고찰에 사용된다.

로직모델에는 시스템 기반 로직모델과 과정 지향 로직모델이 있다. 시스템 기반 로직모델은 시스템의 복잡성(예: 시스템의 지리적, 역학적, 정치적, 사회문화적, 사회경제적 특징)과 맥락적 특징, 연구대상자 및 중재 사이의 상호작용을 조사하는 데 주로 쓰인다. 과정 지향 로직모델은 중재가 중재결과로 이어지는 인과 경로와 중재 효과를 수정할 수 있는 요인의 복잡성을 포착하는 것을 목표로 한다. 그러나, 이들이 명확하게 구별되는 것은 아니며 두 유형은 서로 관련이 있다. 일부 로직모델은 시스템과 프로세스 모형의 요소를 동시에 가진다.

로직모델을 다이어그램으로 표현함으로써 독자들에게 시각적 요약을 제공하고 의사결정자와 중재 수행자를 위한 의사소통 도구가 될 수 있다. 이를 통해 문헌고찰 범위와 PICO에 대해 서로 다른 이해당사자 간의 이해를 공유하는데 도움을 줄 수 있으며, 연구 질문을 개발하고, 중재결과를 구조화 하는 것과 같은 문헌고찰의 결정 사항을 지원하는 데 도움을 줄 수 있다. 로직모델을 통해 중재군과 대조군의 비교 방법, 하위군 분석과 같은 문헌고찰의 PICO 요소 계획과 합성이 어떻게 구조화될 것인지 결정하는 데 사용될 수 있다.

체계적 문헌고찰에서 로직모델은 다음과 같은 경우에 쓰인다(Anderson 등 2011).

- 문헌고찰 질문 정립
- 문헌고찰 주제를 '광범위' 또는 '좁히기' 중 무엇으로 할지 결정
- 중재 구성 요소 식별
- 중재 정의 및 시행
- 관련 연구 적격 기준 확인
- 문헌 검색 전략 수립
- 문헌고찰에 사용된 대리 중재결과의 근거 설명
- 하위군 분석의 필요성 정당화(예: 연령, 성별, 사회-경제적 지위)
- 정책 및 진료에 적절한 문헌고찰이 되도록 함.
- 연구결과의 구조적인 보고
- 위해와 실현 가능성이 중재와 어떻게 연관되어 있는지 설명
- 중재 이론 및 시스템 사고에 기초한 연구결과 해석

중재가 유익한 효과를 찾지 못하였을 때 그 요인이 이론의 실패, 실행의 실패 또는 둘 다 중 어디에 해당하는 지를 고려할 때 로직모델이 유용하게 쓰인다. 실행과 중재 이론을 구분하는 것은 중재가 그 맥락과 상호 작용하는지 또는 어떻게 상호작용하는지 여부를 결정하는 데 도움이 될 수 있다. 이는 중재가 구현되는 방법의 변화를 통해 의도된 중재결과가 바뀔 수 있을지를 설명하는 데 도움이 될 수 있다.

로직모델은 문헌고찰의 개념화 및 구조화 측면에서 프로토콜에 점점 더 많이 사용된다. 로직모델은 사전에 정해질 수 있으며 문헌고찰과정 내내 변경되지 않을 수 있다. 그러나 문헌고찰의 중재결과가 로직모델을 업데이트하는 데 사용될 수 있는 근거와 새로운 이해를 만들 수도 있다. 일부 문헌고찰에서는(새로운) 근거에 기초하여 모형을 수정할 수 있는 보다 단계화된 접근방식을 취하고, 단계화된 로직모델은 종합 계획의 개정을 보고하는 효율적인 방법을 제공할 수 있다.

2.5.2 문헌고찰 질문 변경

문헌고찰을 시작하기 전까지 프로토콜에서 정한 핵심질문을 견지해야 하지만 예기치 않은 문제가 생길 때는 변경을 고려할 수 있다. 일반적으로 문제와 관련된 일부 연구를 알지 못한 상태에서 대답 가능한 질문을 공식화하는 것은 불가능하다. 문헌고찰을 수행하는 과정에서 축적된 근거에 비추어 문헌고찰에서 다루어야 할 질문을 수정할 수도 있다. 다만 질문 수정에 따른 비뚤림을 경계해야 한다. 질문을 조정할 때 다음과 같은 질문을 하는 것이 유용하다.

- 개정의 동기는 무엇인가?
- 포함된 연구결과에 따라 개정작업이 영향을 받을 수 있었는가?
- 세분화된 질문으로 인해 검색 전략의 수정 및 연구 포함 적격에 관한 결정에 재평가가 필요한가?
- 자료 수집 방법이 개정된 질문에 적절한가?
- 개정된 질문이 여전히 앞 절에서 논의된 FINER 기준을 충족하는가?

2.5.3 드문 자료 처리를 위한 비상계획의 수립

문헌고찰 질문을 해결할 수 있는가는 근거의 성숙도와 타당도에 따라 달라진다. 연구가 거의 없으면 정보의 양적 합성을 통해 문제를 해결할 수 있는 기회는 제한적일 것이다. 이런 경우를 대비하여 더 광범위한 그룹화가 되도록 PICO 요소를 불확실하게 남겨놓을 수 있다. 이렇게 하면 잠재적으로 더 많은 수의 연구를 통합할 수 있다. 더 광범위한 그룹화는 일반적으로 다음의 예와 같이 덜 구체적인 문제를 다루는 것이다.

- '비타민 C가 ...에 미치는 영향' 대신 '항산화제 보충제가 ...에 미치는 영향';
- '성감염에 대한 성건강증진의 영향' 대신 '성건강증진이 생물학적 중재결과에 미치는 영향' 또는
- '어린이의 인지 행동 요법이 ...에 미치는 영향' 대신 '어린이와 청소년의 인지 행동 요법이 ...에 미치는 영향'

이러한 광범위한 질문은 효과적인 중재가 부족한 분야에서 중요한 단서를 식별하고 미래 연구를 지도하는 데 유용할 수 있다. 그룹화의 변화는 근거의 확실성 평가에 영향을 미칠 수 있다.

2.5.4 경제성 자료

예를 들어, 특정 중재를 채택하면 자원을 보다 효율적으로 사용할 수 있는지와 같은 질문같이 의사결정자들은 중재의 경제적 측면을 고려할 필요가 있다. 따라서 자원 사용, 비용 또는 비용 효과(또는 이것의 조합)와 같은 경제성 자료는 문헌고찰의 연구결과에 포함될 수 있다. 자원 사용과 비용의 측정치는 특정 항목이나 범주의 수준으로 세분화하는 것이 유용하다.

2.6 주제 범위 관련 MECIR

C1: 문헌고찰 질문을 정하기(필수)	
문헌고찰의 질문 및 특히 관심 중재결과가 소비자, 보건 전문가 및 정책 입안자와 같은 사용자에게 중요한 문제인지 확인한다.	코크란 리뷰는 단지 과학적 호기심이 아니라 임상의료와 정책을 지원하기 위한 것이다. 소비자의 요구는 코크란 리뷰에서 중심적인 역할을 하며 그들은 문헌고찰 질문을 정의하는 데 중요한 역할을 할 수 있다. 질적 연구, 즉 중재를 하거나 받는 관련된 사람들의 경험을 탐구하는 연구와 중재 시행을 구체화하는 요소를 평가하는 연구도 같은 방법으로 사용될 수 있다.
C2: 목표를 사전에 정의하기(필수)	
사전에 연구대상자, 중재, 대조, 중재결과(PICO)를 포함하는 문헌고찰의 목표를 정한다.	목표에 따라 문헌고찰의 초점을 구성하며 이는 적절한 적격 기준을 만들기 전에 명확해야 한다. 복수의 중재를 다루는 경우, 이러한 중재를 어떻게 다룰 것인지(예: 분리해서 요약, 결합하거나 또는 명시적으로 비교)에 대해 명확성이 요구된다.

C3: 잠재적인 위해 효과를 고려하기(필수)	
중재의 잠재적인 중요한 위해 효과를 고려하였고 이것이 언급되었는지 확인한다.	근거의 일방적 요약을 피하기 위해, 위해 효과에 대처하는 것이 중요하다. 최소한 포함된 모든 연구에서 잠재적 위해 효과를 평가한 부분을 강조할 필요가 있다. 때때로 위해 효과에 대한 자료는 무작위배정 되지 않은 연구 혹은 질적 연구에서 가장 잘 얻어진다. 그렇다고 해서 모든 문헌고찰에 비무작위 연구가 반드시 포함되어야 한다는 의미는 아니다.
C4: 형평성과 특정 인구집단에 대한 고려(권고)	
특정 인구집단에 대한 형평성 및 근거의 타당도 문제가 문헌고찰에 중요한지 여부를 미리 고려하고, 문제가 있는 경우 이를 해결하기 위한 적절한 방법을 계획한다. 문헌고찰 질문이 사회경제적 수준이 낮은 집단, 중산층, 여성, 어린이, 노인 등 인구와 관련성이 있는지 주목해야 한다.	가능한 경우, 전체 인구뿐만 아니라 혜택 받지 못한 사람들도 대상으로 포함해야 한다. 그리고 보건의료에서 사회 경제적 불평등을 줄이고 지역사회에서 중재의 사용을 촉진하기 위한 중재의 능력에 대한 명확한 설명을 포함해야 한다.

2.7 참고문헌과 읽을거리

- Anderson L, Petticrew M, Rehfuess E, Armstrong R, Ueffing E, Baker P, Francis D, Tugwell P. Using logic models to capture complexity in systematic reviews. Research Synthesis Methods 2011; 2: 33-42.
- Cummings SR, Browner WS, Hulley SB. Conceiving the research question and developing the study plan. In: Hulley SB, Cummings SR, Browner WS, editors. Designing Clinical Research: An Epidemiological Approach. 4th ed. Philadelphia(PA): Lippincott Williams & Wilkins; 2007. p. 14-22.
- Hoffmann T, Glasziou P, Boutron I. Better reporting of interventions: template for intervention description and replication(TIDieR) checklist and guide. BMJ 2014; 348: g1687.
- Lewin S, Hendry M, Chandler J, Oxman AD, Michie S, Shepperd S, Reeves BC, Tugwell P, Hannes K, Rehfuess EA, Welch V, McKenzie JE, Burford B, Petkovic J, Anderson LM, Harris J, Noyes J. Assessing the complexity of interventions within systematic reviews: development, content and use of a new tool(iCAT_SR). BMC Medical Research Methodology 2017; 17: 76.
- PROSPERO 등록부(https://www.crd.york.ac.uk/prospero/)
- Thomas J, Kneale D, McKenzie JE, Brennan SE, Bhaumik S. Chapter 2: Determining the scope of the review and the questions it will address. In: Higgins JPT, Thomas

J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.0 (updated July 2019). Cochrane, 2019. Available from www.training.cochrane.org/handbook.

제3장 선정기준결정과 PICO 범위 설정하기

핵심 포인트

- 문헌고찰의 범위는 인구집단(연구대상자), 중재 종류(그리고 비교 중재), 관심 있는 중재결과에 의해 정의된다.
- 연구질문의 인구집단, 중재, 비교중재, 그리고 연구의 종류는 사전 적격 기준을 구성할 수 있다. 중재결과를 적격 기준으로 사용하는 경우는 드물다. 중재결과 보고에 관계없이 연구를 포함해야 하지만 중재결과가 측정되지 않았거나, 특정 중재결과를 명백히 보고하지 않으려는 경우 해당 연구는 제외할 수 있다.
- 체계적 문헌고찰에서는 의미 있을 가능성이 높은 모든 중재결과를 포함해야 하며 사소한 중재결과는 포함하지 않아야 한다. 핵심적이고 중요한 중재결과의 개수는 제한해야 하며 이득과 위해 중재결과를 모두 포함해야 한다.
- 문헌고찰 저자는 프로토콜 단계에서 연구의 범위 안에 포함된 다양한 연구대상, 중재, 중재결과 및 연구설계를 분석 단계에서 어떻게 묶을 것인지 계획해야 한다.

3.1 서론

체계적 문헌고찰과 서술적 문헌고찰을 구분하는 특징 중 하나는 체계적 문헌고찰 저자가 문헌고찰에 포함될 연구를 선정하거나 제외하는 기준을 사전에 정해야 한다는 것이다. 프로토콜을 개발할 때 첫 번째 단계는 연구 질문의 요소(인구집단, 중재, 비교중재, 중재결과)와 특정 인구집단에서 중재를 통해 어떻게 기대되는 중재결과가 나타나는지 결정하는 것이다.

적격 기준은 연구질문의 PICO 요소에, 이런 질문을 다루는 연구 유형을 정하여 만든다. 연구 질문에서 인구집단, 중재, 비교중재는 보통 체계적 문헌고찰의 적격 기준으로 바로 전환되기도 하지만, 항상 간단하지는 않으며, 사려 깊게 접근하는 것이 필요하다. 중재결과는 보통 연구를 포함하는 기준으로 간주되지는 않는다.

PICO 형식을 사용할 때 1) 문헌고찰 PICO(연구의 적격 기준의 토대가 됨), 2) 개별 합성을 위한 PICO(각각의 특정 합성이 답하고자 하는 질문을 정의), 3) 포함된 연구의 PICO(포함된 연구에서 실제로 조사된 것)로 구분하는 것이 필요한 의사결정을 이해하는데 도움이 된다. 이 장에서는 문헌고찰 PICO와 개별 합성을 위한 PICO에 초점을 맞추었는데 이것들은 어떤 연구가 포함되어야 하는지를 구체화하고 합성을 계획하는데 기초가 된다. 이 PICO 들은 문헌고찰 질문이나 가설과 명확하고 직접적으로 관련이 있어야 하며, 연구 질문의 대상을 구체화하고

중재군 간의 비교를 포함해야 한다. 문헌고찰을 시작하는 과정에서 중재(예. 약물의 개별 화합물), 인구집단(예. 급성인지 만성인지), 중재결과(예: 여러 우울증 척도) 및 연구설계의 어떤 특성을 함께 그룹화할 것인지 명시해야 한다. 이런 결정은 어떤 연구가 포함될 것인지 그리고 사용될 합성 방법(예: 메타분석)이 무엇인지를 모르는 상태에서 이루어져야 한다.

수집된 자료를 보고 새로운 비교 중재를 추가할 필요가 있을 수 있다. 예를 들어, 자료가 수집된 후에야 중재의 중요한 변화가 발견될 수 있고, 비교를 위한 PICO를 충족하는 연구가 거의 없을 때 비교대상을 수정하면 합성 가능성을 높일 수도 있다. 한편, 체계적 문헌고찰 수행 중 리뷰 프로토콜과 달리 수행 중에 변경된 내용은 최종보고서에 기록되고 문서화되어야 한다.

3.2 문헌고찰과 비교 PICO 명확화하기

3.2.1 인구집단의 유형 정의하기: 어떤 연구대상자와 인구집단?

문헌고찰에서 연구에 포함된 인구집단의 유형을 고려하는 기준은 어떠해야 하는가? 그것은 연구의 다양성 및 중재가 사용될 가능성이 있는 시나리오를 포괄할 수 있을 정도로 충분히 넓어야 하지만, 의미 있는 답을 얻을 수 있을 정도로 충분히 좁아야 한다. 인구집단의 특성이 연구대상자의 수준(예. 나이, 중증도)인지, 연구의 수준(예. 의료 환경, 지리적 위치)인지에 대한 고려가 있어야 하는데, 이는 연구를 그룹화하고 합성할 때 영향을 미치기 때문이다(8장 메타분석 수행). 관심 있는 인구집단의 유형을 고려할 때 다음과 같은 3단계가 도움이 될 것이다.

첫째, **관심 질병이나 건강상태는 그것들의 존재(또는 부재)를 확증하기 위한 명확한 기준**을 사용하여 정의해야 한다. 예를 들어 최근에 개발된 진단 기준은 예전 연구에서는 사용되지 않았을 것이다. 비싸거나 최근에 개발된 진단 검사는 많은 국가나 환경에서 사용할 수 없으며, 일상적인 의료 환경에서는 시간이 많이 걸리는 검사는 실용적이지 않을 수 있다.

둘째, **관심 있는 인구집단과 세팅을 넓게** 정의해야 한다. 여기에는 나이, 성별, 인종, 교육 상태 또는 협심증, 호흡곤란과 같은 특정 상태의 존재 여부로 정의된 인구집단이 구체적으로 어떤 범위에 있는지와 관련된다.

연구대상자를 정의하고 세팅을 정하는데 사용하는 기준이 무엇이든 간에 문헌고찰의 연구대상자와 부분적으로만 겹치는 연구가 발생할 수 있다. 예를 들어 아동에 관한 문헌고찰에서는 16세 미만이 선정기준이라고 한다면, 어떤 연구에서는 12세에서 18세 사이의 연구대상자들이 포함된 경우들이 있다. 이 연구에서 문헌고찰에 적절한 연구대상자를 따로 제시한 경우가 아니라면 저자는 이러한 연구를 다루기 위한 전략이 필요하다.

즉, 적격 기준에 해당하지 않는 연구대상자를 포함함으로써 적용성이 감소하는 것과, 연구를 제외함으로써 자료를 잃는 것 사이에서 균형을 찾아야 한다. '16세 이하의 연구대상자가 80% 이상인 경우 연구를 선정하는'것과 같은 임의적인 규칙은, 연구에서 자세한 정보를 얻을

수 없다면 실용적인 방법은 아닐 것이다.'연구대상자의 대부분이 16세 이하'와 같이 덜 엄격한 규칙만으로 충분할 수 있다. 어려운 결정은 자문 그룹에게 확인하고, 민감도 분석을 해서 이러한 결정이 결과에 미치는 영향을 평가할 수 있다.

'인구집단의 유형'에 대해 기준을 정할 때 고려할 요소는 아래와 같다.

- 질병/건강상태를 어떻게 정의하고 있는가?
- 연구대상자를 설명하는 가장 중요한 특성은 무엇인가?
- 관련 인구학적 요인이 있는가?(예. 나이, 성별, 민족)
- 세팅은 무엇인가?(예. 병원, 지역사회 등)
- 누가 진단을 내리는가?
- 문헌고찰에서 제외되어야 할 다른 유형의 연구대상자가 있는가?(다른 방식으로 중재에 반응할 가능성이 있기 때문에)
- 적절한 연구대상자가 일부만 포함된 연구는 어떻게 처리할 것인가?

셋째, **중재 효과의 크기가 변화될 것으로 예상되는 인구집단 특성**이 있는지를 고려해야 한다(예. 심부전의 다양한 중증도). 하위군을 확인하는 것이 중재를 실제 사용할 때 중요할 수 있다. 적절한 하위군이 밝혀지면, 특정 하위군을 제외하기 위해 문헌고찰 범위를 제한하거나 문헌고찰의 범위를 유지하고 분석에서 하위군을 다루는 방법이 가능하다.

어떤 인구 특성이나 세팅에 대해 문헌고찰을 제한하는 것은 적절한 근거를 바탕으로 해야 한다. 단순히 개인적인 관심으로 나이, 성별 또는 민족성에 따라 특정 하위군에 대해 문헌고찰의 초점을 맞추는 것은, 그렇게 할 만한 생물학적이나 사회학적 정당성이 존재하지 않는 경우, 피해야 하는데, 의사결정권자나 독자들에게 덜 유용할 것이기 때문이다.

문헌고찰의 범위를 유지하는 것은 연구대상자의 다양한 하위군 간에 효과의 중요한 차이가 있는지 불확실할 때 좋을 수 있다. 문헌고찰 저자는 하나의 합성 내에서 다양한 하위군으로부터 얻어진 결과를 합성함으로써 주어진 하위군이 중재 효과간의 변동성(이질성)을 설명할 수 있는지를 보여줄 수 있다. 대안으로, 다른 하위군별로 비교중재를 구별하여 결과들을 합성할 수도 있을 것이다. 하위군별로 분리하는 것은 유용한 합성 결과를 산출하기에는 너무 적은 연구수를 가질 위험이 있다. 특히 연구대상자 수준에서 측정된 특성에 대해서는 하위군 분석 방법에 대한 고려가 필요하다. 모든 하위군 분석은 원칙적으로 미리 계획되어야 하며 프로토콜에서 2차 목표로 명시되어야 하고 자료의 이용 가능성에 의해 계획되어서는 안 된다.

3.2.2 중재 정의 및 그룹화 방법은?

중재를 사전에 정의하는 것이 간단할 수도 있다. 예를 들어 심부 정맥혈전증에 대한 항응고제 효과의 문헌고찰에서는 중재를 정확하게 정의할 수 있다. 하지만 "비만율을 줄이기 위한 식

이 상담, 훈련 등을 포함한 다요소 중재(multicomponent intervention)"와 같은 경우에는 더 복잡한 정의가 필요할 수 있다.

중재를 정의할 때 내재적인 복합성이 나타나는 경우 해결하는 방법은 두 가지이다. 첫째는 의도된 효과가 어떻게 달성되는지를 고려하는 방법이다. 이 방법의 예는 "항응고제인 와파린은 응고 인자를 생성하는 비타민 K에 의존하는 효소를 차단함으로써 혈전을 감소시키는 것으로 보인다"와 같은 방법이다. 두 번째는 변형된 형태의 중재에 따라 효과가 달라질 가능성이 있는지를 고려하는 것이다. 이 방법의 예는 "행동 중재(behavioural intervention)는 건강식을 준비하는 능력에 따라 개인의 자기효능감을 증가시키는 것으로 생각된다"와 같은 것이다. 두 가지 예에서 우리는 모든 형태의 중재가 같은 방식으로 작동한다고 가정할 수는 없다는 것을 알 수 있다. 항응고제와 같은 약물 중재를 정의할 때는 약물 준비, 투여 경로, 용량, 기간 및 빈도와 같은 요소를 고려해야 한다. 비만율을 낮추기 위한 중재 같이 다요소 중재의 경우 중재의 공통 또는 핵심 특성을 정의해야 문헌고찰에 포함되지 않은 다른 중재와 명확하게 구분할 수 있다.

일반적으로 무엇을, 누가, 어떻게, 어디에서, 언제, 얼마만큼 투여할지, 그리고 그 중재를 조정하거나 환자 맞춤으로 수정할 수 있는지 여부를 정확하게 고려하는 것이 유용하고, 이것을 문헌고찰에서 포함된 중재의 각 유형별로 고려하는 것이 유용하다(TIDieR checklist 참고). 중재를 '단순'과 '복합'으로 나누는 것은 잘못된 이분법이다. 모든 중재는 어떤 면에서는 복잡할 수 있다. 문헌고찰 저자에게 중요한 문제는 특정 문헌고찰에서 고려해야 할 가장 중요한 요소를 정하는 것이다.

중재의 유형'에 대한 기준을 개발할 때 고려할 요소는 아래와 같다.

- 관심 있는 실험 중재와 대조 중재는 무엇인가?
- 중재가 다양한가?(예. 용량/강도, 투여 방법, 투여 담당자, 빈도, 투여 시간 등)
- 모든 다양성을 포함해야 하는가?(예를 들어, 임상적으로 적절하지 않은 용량을 모두 포함할 것인가)
- 중재의 일부만을 포함하는 연구를 선정할 것인가?
- 다른 중재와 결합된 중재를 포함할 것인가?
- '대조', '위약', '무 치료', '일반적 관리'를 다른 의미로 고려할 것인가?

일단 문헌고찰 대상으로 정한 중재가 광범위하게 정의되면, 중재의 변이를 합성 과정에서 어떻게 처리할지 결정해야 한다. 연구 간에 중재 특성의 차이는 모든 문헌고찰에서 나타난다. 간단한 음주 상담 중재의 기간이나 내용의 사소한 차이 같이 실제 사용되는 중재방식의 사소한 차이를 반영한다면 이를 전체적으로 합성한 결과를 제시하는 것이 의사결정자에게 유용한 정보를 제공할 것이다. 중재 특성의 차이가 더 크고(예. 간호사 vs. 의사의 간단한 음주 상담 중재) 중재 효과의 크기의 차이가 클 것이라고 예상되는 경우 이러한 차이는 합성 중에 고려되어야 한다.

명확하게 중재를 그룹화하면 합성에서 중요한 두 가지 역할을 한다. 첫째, 중재가 합성(메타분석 혹은 다른 합성)을 위해 그룹화 되는 방법에 따라 문헌고찰의 결과에 영향을 미칠 수 있다. 둘째, 프로토콜에 명시된 중재군은 연구 저자들이 보고서에 사용한 다양한 설명을 극복할 수 있게 해주고, 문헌고찰 전반에 걸친 중재를 설명하기 위한 표준화된 용어를 제공한다. 이 표준화는 연구들 사이 중재의 특성에 대한 정보의 비교 및 합성을 가능하게 하고 문헌고찰 결과의 해석을 뒷받침하는 보고를 위한 일관된 언어를 제공한다.

3.2.3 비교 중재 정의하기

합성을 위한 PICO를 표현할 때, 중재군만 정의하는 것은 합성 계획으로 충분하지 않다. 다음 단계는 중재군 간에 비교가 이루어지도록 정의하는 것이다. 다른 그룹을 동시에 비교할 수 있는 네트워크 메타분석과 달리 표준 메타분석은 한 번에 두 그룹의 비교 효과에 대한 결론을 이끌어 내는 것을 목표로 한다.(즉, 두 중재군 중 어떤 것이 더 효과적인가?). 이러한 비교는 자료를 이용할 수 있을 경우 기본적으로 하는 합성 방법이다.

비교의 유형은 다음 세 가지이다.

- 중재 vs. 위약(예. 위약, 가짜 수술, 심리적 위약). 위약은 약리학적 중재를 평가할 때 가장 일반적으로 사용되지만 일부 비약리학적 평가에서도 사용될 수 있다. 예를 들어,
  - ✓ 신세대 항우울제 vs. 위약
  - ✓ 골다공증성 척추 압박골절에 대한 척추 성형술 vs. 플라시보(가짜수술)
  - ✓ 중재 vs. 대조군(예. 중재 없음, 대기자 대조군, 일반적인 치료). 두 치료군 다 표준 치료를 받을 수 있다. 예를 들어,
  - ✓ 식도 및 위식도 접합부위 암 환자를 위한 화학요법 또는 표적요법 + 최상의 지지 치료 vs. 완화치료를 위한 최상의 지지 치료
  - ✓ 장기간 건강문제를 가지고 있는 사람들에게 개인 맞춤형 치료 계획 vs. 일상적인 치료

- 중재 A vs. 중재 B. 대조군으로서 같은 중재를 사용하면서 다른 시각, 다른 기간이나 다른 용량, 혹은 두 개의 다른 중재를 포함할 수 있다. 예를 들어,
  - ✓ 만삭아와 조산아에서 비경구 아연 보충제를 공급하는데 있어 초기(생후 2주 미만) vs. 후기(생후 2주 이후)
  - ✓ 고관절염 혹은 슬관절염 환자에서 고강도 vs. 저강도 신체활동이나 운동
  - ✓ 처방 의약품과 일반 의약품에 대해 소비자를 상대로 하는 멀티미디어 교육 vs. 다른 교육

3.2.3.1 공통 중재 다루기

동일한 중재가 시험군과 대조군에 동시에 제공되는 특수한 경우에 대해서는 계획이 필요하다. 공통 중재는 추가적인 중재로 관심 있는 중재와 함께 제공되는 것으로 예를 들어 아로마 테라피의 효과를 보고자 하는 문헌고찰에서 마사지 같은 것이다(즉 아로마 테라피+마사지 vs. 마사지). 많은 경우 공통 중재는 덜 중요한 것으로 여기고 무시될 수 있다. 하지만 다른 상황에서는 연관 있는 중재의 효과가 공통 중재의 유무에 따라 다르게 나타날 수도 있다. 예를 들어 마사지를 받으면서 나타나는 아로마 테라피의 효과와 아로마 테라피만 받는 경우의 효과가 다를 수 있다. 이것은 관심 있는 중재와 공통 중재 간에 상호작용이 있어서 효과를 증대시키거나 아니면 낮추거나 하기 때문이다. 정성적 상호작용은 드물지만(공통 중재와 결합될 때 중재의 효과가 반대방향으로 나타나는 경우), 공통 중재가 있을 때 중재 효과가 더 다양하게(이질성) 나타날 수 있으며, 이를 어떻게 처리할지에 대한 계획을 수립 하는 것이 중요하다. 이 문제를 통계학적 합성 때 다루는 방법으로는 변량효과모형(random effect model)을 적용하거나, 중재 효과가 공통 중재에 따라 바뀐다면 하위군 분석을 통해서 해결하거나 하는 방법이 있다.

3.2.4 중재결과를 선택하고 우선순위를 부여하고 그룹화 하기

3.2.4.1 중재결과를 선택하기

넓은 범위의 중재결과 영역은 문헌고찰 PICO를 정할 때 결정된다. 넓은 영역이 합의되면, 추가적으로 보고와 합성을 용이하게 하기 위해 영역을 정의하는 것이 필요하다(즉 비교를 위한 PICO). 중재결과를 구체화하고 그룹화하는 프로세스는 중재군을 정하기 위해 사용되는 것과 유사하다.

중재결과에 대한 보고를 문헌고찰을 위한 적격 기준으로 삼아서는 안 된다. 특히 연구에서 측정한 결과를 보고하지 않거나 '쓸모없는 자료'를 제공하였다고 해서 연구를 제외하면 안 된다. 이는 연구 저자가 중재결과를 선택적으로 보고하면서 생길 수 있는 비뚤림을 피하기 위해 필수적이다. 그러나 어떤 경우에는 특정 중재결과의 측정이 연구 적격 기준이 될 수 있다. 예를 들어 문헌고찰에서 특정 중재결과를 예방하기 위한 중재의 가능성을 다루거나, 문헌고찰에서 같은 연구 대상에게 다른 목적으로 사용하는 경우 중재의 특정 목적을 다룰 때(호르몬 대체요법, 아스피린 같은)가 이에 해당될 수 있다.

일반적으로 의도된 사용자와 검토된 근거를 받아들이는 사람들에게 의미가 있을 것 같은 중재결과를 포함시키도록 해야 한다. 여기에는 임상 의사, 환자(소비자), 일반 대중, 관리자 및 정책 결정자가 포함될 수 있다.

의사결정자에게 사소하거나 의미가 없는 중재결과는 체계적 문헌고찰에 포함되어서는 안 된다. 중요성이 거의 또는 전혀 없는 중재결과를 포함하면 독자들이 오해할 수 있다. 의미 있는 것으로 여겨져서 문헌고찰에서 다루기로 한 중재결과가 일차 연구에서 보고되지 않았을 수도 있다. 예를 들어, 삶의 질은 중요한 중재결과이고, 진행 암에 대한 화학요법의 사용여부를 결정하는데 어쩌면 가장 중요한 중재결과인데, 이용 가능한 연구들에서는 오로지 생존 결과만

보고하는 경우가 있다. 모든 중요한 중재결과를 문헌고찰에 포함시키고 논의하는 것을 통해 일차 연구에서의 간극을 강조하고 연구자가 앞으로 진행될 연구에서 이러한 간극을 다루도록 장려할 수 있다.

3.2.4.2 중재결과의 우선순위 설정하기

적절한 중재결과 목록이 수집되면 저자는 우선순위를 설정하여 문헌고찰 질문에 가장 적절한 중재결과를 선택해야 한다. GRADE에서는 문헌고찰 저자가 의사결정을 위해 중재결과 를 '핵심적', '중요한', '중요하지 않은'으로 나누어 제시하도록 한다.

핵심적 중재결과는 의사결정에 필수적인 결과이며 '결과 요약표(Summary of findings)'나 다른 요약표의 기초를 이룬다. 결과 요약표는 중요한 비교중재와 중재결과에 대한 근거의 양과 근거의 질 및 효과의 크기에 대한 핵심 정보를 제공한다.

중재결과를 선택하고 우선순위를 설정할 때 고려할 요소는 아래와 같다.

- 모든 잠재적인 의사결정자와 연관된 중재결과를 고려한다.
- 핵심적 중재결과는 의사결정에서 필수적인 것이다, 환자에게 중요한 중재결과에 중점을 두며, 핵심결과 세트(core outcomes sets)에 의해 결정되어야 한다.
- 의사결정자에게 중요한 추가적인 중재결과도 문헌고찰에 포함될 수 있다. 의사 결정자에게 중요하지 않은 중재결과는 문헌고찰에서 제외해야 한다.
- 최대 7개의 핵심적이고 중요한 중재결과를 선택해야 하여, 이 내용을 결과 요약표, 초록 및 일상 언어 요약(plane language summary)에 제시한다. 요약문만 단독으로 읽을 수 있기 때문에 의사 결정자에게 가장 중요한 중재결과를 포함시키는 것이 중요하다.
- 중재결과에 잠재적일뿐만 아니라 실제적인 유해효과가 포함되어 있는지 확인한다.

3.2.4.3 합성을 위해 중재결과를 정의하고 그룹화 하기

동일한 임상영역에서 시행한 임상시험 간에도 측정된 중재결과가 불일치할 수 있다. 이러한 불일치는 고려되는 영역의 폭, 측정된 중재결과, 그리고 이러한 결과를 명명하고 정의하는 방식, 측정 방법과 시기 등 결과 측정의 모든 측면에 나타난다. COMET(Core Outcome Measures in Effectiveness Trials)는 임상시험 간 중재결과 측정의 표준화를 권장하는 것을 목표로 하고 있지만, 이러한 이니셔티브는 비교적 새로운 것이고, 문헌고찰 저자들은 연구 간 중재결과의 다양성을 만나게 될 것이다. 다음의 순서로 합성을 위한 중재결과를 정의하고 그룹화 할 수 있다.

1. 중재결과 영역을 명확히 결정한다.

각 중재결과 영역에 대해 짧은 라벨(예. 인지, 진료에 대한 소비자 평가)을 부여하고 포함된 각 연구의 적절한 중재결과를 분류할 수 있도록 영역을 충분히 자세히 설명한다. 정의는 측정된 개념(또는 구성), 즉 '무엇'이 측정되는지를 기반으로 해야 한다. '언제' 그리고 '어떤' 중재

결과가 측정될 것인지는 다음 단계에서 고려한다.

중재결과는 매우 광범위한 그룹(예. 생리/임상적 결과, 삶에 미치는 영향, 유해효과)에서 시작해서 중재결과 영역(예. '기능' 및 '인지된 건강상태'는 '삶에 미치는 영향'의 영역이 됨)으로 계층적으로 정의될 수 있다.

보다 좁은 영역(예. 신체 기능, 인지 기능)내에서 구체적인 중재결과가 측정된다. 합성에서 어떤 결과의 수준이 그룹화 될지에 따라서 다루는 연구 질문이 변화 될 것이므로, 문헌고찰 목적에 의해 결정되어야 한다.

중재결과 영역을 결정할 때

- 가능하면 정의는 존재하는 기존 시스템을 반영하거나 의사 결정자가 이해할 수 있는 적절한 문헌과 용어를 반영해야 함
- 중재결과가 불일치하게 표시되고 설명되어야 할 경우, 목록의 예시를 보여주면 영역의 범위를 전달할 수 있음
- 영역이 정의되는 수준(넓게 vs. 좁게)과 보고 및 합성에의 적용을 고려할 것: 다양한 중재결과를 결합하면 설명할 수 없는 이질성이 생길 수 있는 반면, 좁게 정해진 중재결과는 보고한 연구가 거의 없으면 합성이 어려울 수 있음
- 인과 관계도 또는 로직 모델은 보고 및 분석을 위한 관련 중재결과의 논리적 그룹화 및 합성할 대체 수준을 알아내는데 도움을 줄 수 있음

2. 중요한 중재결과를 식별하고 그룹화하기 위한 기존 시스템이 있는지 결정한다.

중재결과를 분류하는 시스템으로는 COMET, ICHOM 이니셔티브 및 결과 분류법과 같은 핵심 중재결과 세트(core outcome set)가 있다. 이 시스템은 연구 전반에 걸쳐 사용되는 다양한 중재결과의 라벨을 표준화 하고 그룹화 및 비교를 가능하게 하는데 사용될 수 있다. 합의된 용어는 의사 결정자가 문헌고찰의 결과를 해석하는데 도움을 줄 수 있다.

COMET 웹 사이트는 합의되거나 개발 중인 핵심 중재결과 세트에 대한 데이터베이스를 제공한다. 일부 코크란 그룹은 자체적으로 중재결과 세트를 개발하였다. 중재결과 세트와 분류법은 임상영역에 따라 다르긴 하지만, 넓은 결과 영역을 정하기 위한 다양한 분류법이 존재한다.

3. 중재결과 측정 시점을 정의한다.

중재결과를 정의하는 핵심 속성은 측정 시점을 정하는 것이다. 문헌고찰에서 특정 시점이 아니라 시간 프레임을 사용하는 것은 종종 연구들 사이의 중재결과 측정 시점이 제각각인 것을 다루기 위함이다.(예. 중기는 6개월부터 12개월 사이에 측정된 중재결과를 포함하는 것으로 정의 될 수 있음)

중재결과 시점을 정할 때

- '임상적으로 의미 있는'시점에 초점을 맞출 것
- 합의 또는 승인된 결과시점(예. 만성 요통에 대한 추적 관찰기간으로 NIH TF에서 제시한 6개월에서 12개월을 사용하는 것), 또는 핵심 중재결과 세트가 있는지 고려할 것
- 시간 프레임의 범위를 신중히 고려할 것. 좁은 시간 프레임은 합성할 때 해당 연구가 거의 없을 수 있다. 넓은 시간 프레임은 연구들 마다 시점이 매우 다양하다면, (지나친) 다양성(Step 5) 때문에 해석하기 어려울 수 있다.

4. 측정 도구나 방법을 정한다.

- 영역이나 특정 중재결과를 평가하기에 적절한 측정 방법이나 도구(예. 임상 평가, 실험실 검사, 객관적 측정, 환자보고 중재결과 등)
- 영역을 측정하는 다른 방법이나 도구가 있는지 여부(Step 6)
- 신뢰도와 타당도의 근거가 있는 것
- 자가보고 측정 도구에서 결과/영역을 다루고 있는 문항들은 어떤 이론이나 경험적 근거, 소비자 개입을 통해 개발된 것임.

측정은 중재결과 세트(예. Williamson et al(2017), ICHOM(2018))에서 가져오거나 도구에 대한 체계적 문헌고찰(COSMIN[1])을 통해 결정할 수 있다.

5. 중재결과의 다양성(multiplicity)을 어떻게 다룰 것인지 정한다.

어떤 영역은 한 연구 내에 여러 중재결과를 포함할 수 있는데 다음과 같은 경우이다.

- 영역 내에서 측정된 여러 중재결과 변수(예. '정신 건강' 영역의 '불안', '우울')
- 중재결과를 측정하기 위한 다양한 방법(예. 자기 보고로 보고된 우울, 임상의가 평가한 우울) 또는 다양한 임상심리측정도구(예. HAM-D, BDI).
- 시간 프레임 내에서 여러 시점에서 측정함.

이렇게 다양하게 측정되어 계산된 중재의 효과는 동일한 연구대상자에게서 계산되었기 때문에 통계적으로 독립적이지 않다. 이러한 의존성을 처리하려면 어떤 비교에 대해 연구마다 하나의 중재결과만을 선택하거나 의존성을 설명하는 메타분석 방법을 사용한다(Step 6).

연구결과와 독립적인 접근법을 사용하여 프로토콜에서 여러 결과나 측정 항목 중 선택하는 방법을 미리 정한다. '포함된 연구의 특성' 표에 모든 적합한 결과 변수나 측정 방식을 기록하고, 어떤 중재결과가 선택되었는지와 그 이유도 같이 남겨야한다.

프로토콜을 개발할 때는 모든 형태의 다양성을 예상하기 어려울 수 있다. 중재결과를 선택하기 위해 사용된 사후 접근법은 '프로토콜과 문헌고찰의 차이' 섹션에 기록되어야 한다.

6. 정해진 중재결과 영역을 합성에서 어떻게 사용할 것인지 계획한다. 연구마다 다른 측정방

1) COSMIN: COnsensus-based Standards for the selection of health Measurement INstruments

법이나 도구가 사용된 경우 이들을 어떻게 합성할지 고려해야 한다. 아래와 같은 옵션이 있다:

- 동일한 중재결과(혹은 결과 영역)에 대한 서로 다른 측정값을 함께 합성하라. 이 접근법은 합성 가능성을 극대화 할 것이다. 하위군 또는 민감도 분석을 통해 효과가 측정방법이나 도구의 유형에 따라 변경되었는지, 혹은 견고한지를 확인할 수 있다. 이질성이 증가될 수 있고 이로 인해 변량효과모형을 사용하게 된다.
- 각 중재결과 측정을 개별적으로 합성하라.(예. BDI와 HAM-D 각각 메타분석) 그러나 측정방법이 모두 동일한 영역을 측정하는 경우, 메타분석의 수가 많아지면서 해석상 어려움이 생기고, 1종 오류가 증가될 수 있다.
- 의존성에 대해 모델링하거나 설명하는 메타분석 방법을 사용하여 가능한 효과 평가를 모두 포함하라. 이 옵션은 모든 정보를 사용함으로써 중재 효과를 추정하는데 더 정확할 수 있다는 이점이 있다. 선택사항으로 다변량 메타분석(multivariate MA), 다수준 모형(multilevel model), robust variance estimation 등이 있다.

7. 가능하다면, 좁은 중재결과 영역과 넓은 결과 영역 둘 다 제시하여 다양한 상황을 보여준다. 다양한 수준이나 시간 간격으로 결과를 그룹화 하여 유연하게 제시하는 것을 고려하라. 유연하지 않은 접근법은 합성의 가능성을 약화시킬 수 있으며 특히 연구가 거의 없거나 중재결과가 정의되고 측정되는 방법과 측정 시점이 매우 다양할 경우 더욱 그러하다. 좁은 영역을 사용함으로써 의미 있는 합성을 하기에 자료가 부족한 경우, 보다 넓은 영역이 사용될 수 있다.

3.2.5 어떤 연구설계가 포함될 것인지 결정하기

문헌고찰 저자는 어떤 연구설계가 문헌고찰의 목적을 이루는데 신뢰할만한 자료를 제공할 가능성이 있는지 먼저 고려할 필요가 있다.

3.2.5.1 무작위배정 임상시험연구 포함하기

체계적 문헌고찰은 의료기술의 효과에 대한 질문을 다루기 때문에 주로 무작위배정 임상시험연구에 초점을 맞추고 있고, 무작위배정 임상시험연구가 관심 중재에 대해 적합하다면 포함되어야 한다. 무작위배정은 알려지거나 알려지지 않은(측정되지 않은) 교란요인에 대해 다른 중재군간 연구대상자의 기저 특성의 체계적인 차이점을 예방할 수 있는 유일한 방법이며, 원인과 효과에 대해 다른 어떤 연구 형태보다 더 신뢰성 있는 결과라고 주장할 수 있다. 임상 중재를 누가 받고, 누가 받지 않는 지는 예후 인자를 비롯한 많은 요인에 의해 영향을 받는다.

실용적으로 고려해보면 많은 체계적 문헌고찰이 무작위배정 임상시험연구로 연구설계를 제한하게 된다. 최근 수십 년 동안 무작위배정 임상시험연구를 색인하고 식별하기 위한 인프라를 구축하는데 국제적으로 상당한 투자가 있었다.

무작위배정 임상시험연구로 제한한다 하더라도 적격 기준을 정의할 때는 연구설계의 요소를 고려해야 한다. 예를 들어 군집 무작위배정 임상시험연구나 교차 임상시험이 자격이 있는지, 위약 대조군의 사용, 배정 은폐를 고려할 것인지, 추적기간의 최소한의 추적기간과 같은

기타 기준들도 고려할 것인가와 같은 것들이다. 제한적인 연구설계 기준(비뚤림의 위험은 낮지만 매우 적은 연구가 포함될 것임)과 보다 자유로운 설계 기준(더 많은 연구를 포함시키지만 비뚤림위험이 높을 것임) 사이에는 항상 절충이 필요하다. 더구나 지나치게 광범위한 기준은 오해의 소지가 있는 근거를 포함시킬 수 있을 것이다. 예를 들어 만성 질환 환자의 생존을 개선시키는 치료법에 초점을 맞추는 경우, 연구 질문을 해결 할 수 없다는 점을 분명히 명시하지 않는 한, 매우 짧은 기간 동안 수행한 연구를 검토하는 것은 부적절할 수 있다.

3.2.5.2 비무작위 연구 포함하기

비무작위 연구(그리고 다른 유형)가 포함될 것인지 결정하는 것은 문헌고찰 PICO 수립과 함께 결정된다. 비무작위 연구가 포함될 수 있는 이유로는 다음과 같은 것들이 있다: 1) 무작위배정 임상시험이 위해나 장기적인 결과에 대한 중재의 효과를 다루기에 불가능하거나 특정한 연구 대상이나 세팅에서 수행될 수 없을 때, 2) 무작위배정을 할 수 없는 중재(예. 단일 혹은 소수의 관할 지역에 도입된 정책 변경의 효과).

비무작위 연구설계는 배정에 무작위배정을 사용하지 않는 공통점을 갖고 있는데, 이러한 설계는 비뚤림의 영향을 받기 쉽다는 것을 의미한다. 적격 기준은 명확한 연구설계 특성에 기반해야 하며, 일관성 없이 사용되는 연구 라벨에 의지해서는 안 된다.

비무작위 연구가 포함되는 경우 문헌고찰 저자는 연구를 어떻게 그룹화 하여 합성할 것인지 고려해야 한다. 코크란 비무작위연구 방법론 그룹에서 만든 연구 디자인 특징 분류법이 도움이 될 것이다.

연구설계를 그룹화 하는 것에 대한 결정이 내려지면, 이들을 어떻게 합성할 것인지 계획을 세우는 것이 필요하다. 문헌고찰 저자는 비무작위 연구의 결과를 합성하는 것이 유용한지 여부를 결정해야 하며, 만일 그렇다면 연구설계가 이질성을 설명하는지 검토하기 위해 무작위배정 임상시험과 비무작위 연구의 결과를 같은 합성에 포함시켜야 하는지 아니면 별도의 비교로 합성해야하는지를 결정해야 한다.

3.2.6 출판 상태와 언어에 대한 적절성

일반적으로 모든 유형의 출판물에 보고된 연구를 포함시키는 전략이 비뚤림을 줄일 것이다. 출판되지 않은 연구나 부분적으로 출판된 연구, '회색'문헌 같이 출판 상태에 따라 연구를 제외한다면 이에 대한 강력한 논의가 필요하다. 미발표 연구를 얻는데 생기는 추가적인 어려움을 고려할 때, 문헌고찰에서 확인된 미출판된 연구는 모든 미출판된 연구를 대표하지 못할 수 있다. 하지만 보고 비뚤림에 대해 알려진 것을 고려한다면, 이러한 비뚤림의 발생은 모든 미출판된 연구를 제외함으로써 나타나는 비뚤림보다는 덜 우려된다.

마찬가지로, 모든 언어의 연구를 찾고 분석하는 것은 매우 자원-집약적일 수 있지만, 문헌고찰 저자는 하나의 특정 언어(일반적으로 영어)로 연구를 제한한다면 비뚤림(및 형평성)에 대한 영향을 신중하게 고려해야 한다.

3.3 선정 기준 관련 MECIR

C5: 연구대상을 사전에 명확히 정의하기(필수)	
연구대상자에 대한 적격기준을 미리 정의한다.	사전에 정의되고 모호하지 않은 적격 기준은 SR을 위한 기본 전제조건이다. 문헌고찰에서 연구에 포함된 사람들의 유형을 고려하는 기준은 연구의 다양성을 포괄할 만큼 충분히 넓어야 하지만, 의미 있는 답을 얻을 수 있도록 충분이 좁아야 한다. 연구대상자를 설정할 때 고려해야 하는 사항에는 세팅, 질병 진단이나 건강상태의 정의, 인구학적 요인을 포함한다. 연구대상을 제한할 때는 정확한 근거에 기초해야 한다.
C6: 적격한 연구대상자의 하위 집단을 포함한 연구에 대한 전략을 사전에 정의하기(권고)	
적절한 연구대상자가 하위집단에만 포함되어 있을 경우 어떻게 할 것인지 정의한다.	때로는 하나의 연구에 '적절한' 연구대상자와 '적절치 않은' 연구대상자가 같이 포함되어 있을 경우가 있다. 만약 적절한 연구대상자의 자료를 따로 추출해낼 수 없다면 이러한 상황을 어떻게 처리해야 할지를 미리 정해야 한다.
C13: 적격 기준 변경하기(필수)	
적격 기준이나 중재결과를 변경하는 경우 그 이유를 정당화하라. 특히 연구의 선정 및 배제에 대해 사후 결정을 할 때는, 임의적인 규칙보다는 문헌고찰의 목적을 잘 유지해야 한다.	사전에 적격 기준을 정하는 것은 체계적 문헌고찰의 기본적 특성이다. 그렇지만 예기치 않은 문제가 발생할 수 있다. 저자들은 사후적으로 연구를 제외할 때 현명한 의사결정을 내릴 수 있어야 하며, 가능하다면 민감도 분석을 시행하고 문헌고찰에 문서화해야한다. 프로토콜의 변경은 연구결과 또는 합성 결과에 따라서 이루어져서는 안 되는데, 이는 비뚤림을 초래할 수 있기 때문이다.
C7: 중재와 비교중재에 대해 명확한 기준을 사전에 정의하기(필수)	
포함 연구에서 적격 중재와 이와 비교될 수 있는 비교중재를 미리 정의하라.	사전에 정의된, 모호하지 않은 적격 기준은 체계적 문헌고찰을 위한 기본 전제조건이다. 비교중재를 결정할 때는 특히 명확성이 요구된다. 실험 중재가 비활성 대조 중재(예. 위약, 치료없음, 표준 치료, 혹은 대기 비교)와 비교되는지, 혹은 활성 대조 중재(예. 같은 중재이나 다른 형태인 것, 다른 약물, 다른 종류의 치료)와 비교되는지? 예를 들어, 투여 방법, 용량, 기간, 강도, 동반중재 및 복합중재의 특징과 같은 중재와 비교중재에 대한 어떠한 제한 사항도 사전에 정의되고 설명되어야 한다.
C8: 중재결과의 역할을 명확하게 하기(필수)	
'해당 문헌고찰에 대한 연구를 고려하는 기준' 항목에 나열된 중재결과가(어떤 연구가 포함되었는지에 대한 관심 중재결과의 목록으로서가 아니라) 연구포함기준으로 사용되는지를 명확히 한다.	중재결과 측정은 항상 문헌고찰에서 연구를 포함시키는 기준의 일부로 여겨져서는 안 된다. 그러나 일부 문헌고찰에서는 특정 중재결과를 합법적으로 제한하기도 한다. 예를 들어, 동일한 중재가 동일한 집단에서 상이한 목적으로 연구될 수 있다(예. 호르몬 대체 요법 또는 아스피린). 또는 어떤 문헌고찰은 여러 조건에서 사용된 중재의 유해효과를 구체적으로 다룰 수 있다. 저자가 중재결과에 근거해서 연구를 제외한다면, 적절한 중재결과가 단순히 보고되지 않아서라기보다는 측정되지 않아서 이용가능하지 않음을 확실히 해야 한다.

C14: 중재결과 영역을 사전에 정의하기(필수)	
문헌고찰에 핵심 중재결과와 추가적으로 중요한 중재결과를 미리 정의한다.	중재결과를 특정한다는 것은 중재결과의 영역(예. 삶의 질)과 중재결과의 측정도구(예. SF-36)을 모두 고려한다는 것이다. 중재결과를 미리 정의하는 것은 선택적 결과보고의 위험을 줄여준다. 핵심 중재결과는 가능한 적어야 하고 적어도 하나의 잠재적인 이득과 위해 영역을 고려해야 한다. 적격한 연구들이 확인되면, 이러한 중재결과들을 합성할 수 있어야 하며, 문헌고찰의 결론은 주로 이러한 중재결과에 대한 중재의 영향에 근거할 것이다. 추가로 중요한 중재결과도 구체화될 수 있다. 문헌고찰에서 7개 이상의 중재결과를 측정할 수 있지만, 최대 7개의 핵심 및 중요한 중재결과가 GRADE 평가의 기초가 되며 문헌고찰의 초록(abstract)과 요약(summary)으로 정리된다.
C15: 중재결과 선정하기(필수)	
보건의료 소비자, 보건의료전문가 및 정책결정자와 같은 문헌고찰의 사용자에게 핵심적이거나 중요한 중재결과만을 선택한다.	코크란 리뷰는 임상 진료 및 정책을 지원하기 위한 것으로 소비자에게 핵심적이거나 중요한 중재결과를 다루어야 한다. 이는 프로토콜 단계에서 정해야 한다. 가능하면 핵심 중재결과 세트가 확립되어야 한다. 가능한 경우, 환자보고 중재결과가 포함되어야 한다. 자원 사용 및 비용의 근거가 중재를 적용하거나 대체 관리 전략을 채택하는 의사결정에서 중요한 요소가 되는지를 판단하는 것도 중요하다. 많은 수의 중재결과가 필요하기도 하지만, 문헌고찰을 초점이 맞지 않게 하고, 관리하기 어렵게 만들며, 선택적 중재결과 보고 비뚤림을 일으키기 쉽게 한다. 생화학적, 중간 및 과정 중재결과가 의사결정자에게 중요하다면 고려되어야 한다. 핵심적이거나 중요한 것으로 간주되지 않는 중재결과는 문헌고찰에서 제외될 수 있다.
C16: 중재결과 측정을 미리 정의하기(권고)	
수용 가능한 중재결과 측정 방법이 무엇인지에 대한 세부 사항(진단기준, 척도, 복합 중재결과)을 미리 정의한다.	문헌고찰에서 어떤 중재결과가 필요한지 결정한 후에 저자는 이러한 중재결과를 측정할 수 있는 수용 가능한 방법을 명확히 해야 한다. 그러나 유해효과를 미리 정의하는 것은 어려울 수 있다.
C9: 연구설계를 먼저 정의하기(필수)	
연구설계 라벨보다는 연구설계의 특성에 초점을 맞춰서, 명확하고 모호하지 않은 방법으로 적격 기준을 미리 정의한다.	사전에 정의되고 모호하지 않은 적격 기준은 체계적 문헌고찰을 위한 기본 전제 조건이다. 이것은 비무작위 연구를 고려할 때 특히 중요하다. 연구설계를 정의하기 위해 사용하는 일부 라벨은 모호할 수 있다. 예를 들어 '이중 맹검' 연구는 누가 눈가림이 되었는지가 명확하지 않다. '환자 대조군' 연구는 코호트 안에서 이루어 진 것일 수도 있고 단면연구의 일환으로 수행된 것일 수도 있다. '전향적' 연구는 단지 몇몇 요소들만 전향적으로 정의되거나 수행된 것일 수도 있다.
C11. 연구설계 선정에 대해 정당성을 부여하기(필수)	
특정 연구설계를 선택한 이유를 설명한다.	무작위배정 임상시험만으로 일부 중재 혹은 일부 중재결과를 다루기 어려울 수 있다. 저자는 무작위배정 임상시험으로 문헌고찰을 제한하거나 비무작위 연구를 포함하도록 선택한 이유를 설명할 수 있어야 한다. 선정된 연구설계가 문헌고찰 질문에 적합한지, 잠재적인 비뚤림은 없는지에 대해 설명할 수 있어야 한다.

C10: 무작위배정 비교임상시험 포함하기(필수)	
관심 있는 중재와 중재결과를 평가하는 것이 가능한 경우 무작위배정 비교임상시험을 적격 기준에 포함시킨다.	무작위배정 비교 임상시험은 중재의 효능을 평가하기 위해 가장 좋은 연구설계이다. 문헌고찰에서 다루는 연구 질문을 평가하기 위해 할 수만 있다면 포함시켜야 한다. 그러나 추적조사 기간과 같은 적절한 제외기준을 적용할 수 있다.
C12: 출판 상태에 따라 연구 제외하기(필수)	
제외하는 것이 명확하게 정당화 되지 않는 한, 출판 상태와 상관없이 연구를 포함한다.	미출판 연구(회색 문헌 포함)의 자료를 확보하여 포함하면 출판바뚤림의 영향을 줄일 수 있다. 그러나 확인할 수 있는 미출판 연구가 모든 미출판 연구를 대표할 수 있는 것은 아니다.

3.4 참고문헌과 읽을거리

- Joanne E McKenzie, Sue E Brennan, Rebecca E Ryan, Hilary J Thomson, Renea V Johnston, James Thomas. Chapter 3: Defining the criteria for including studies and how they will be grouped for the synthesis: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA(editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.0. Cochrane, 2019. Available from https://training.cochrane.org/handbook/current/chapter-03
- COMET(Core Outcome Measures in Effectiveness Trials) Initiative. Available from http://www.comet-initiative.org/
- ICHOM – International Consortium for Health Outcomes Measurement. https://www.ichom.org/
- COSMIN tool. Available from https://www.cosmin.nl/cosmin-tools/
- Better reporting of interventions: template for intervention description and replication(TIDieR) checklist and guide. https://www.equator-network.org/reporting-guidelines/tidier/
- McKenzie JE, Brennan SE, Ryan RE, Thomson HJ, Johnston RV, Thomas J. Chapter 3: Defining the criteria for including studies and how they will be grouped for the synthesis. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.0 (updated July 2019). Cochrane, 2019. Available from www.training.cochrane.org/handbook.
- Schünemann HJ, Tugwell P3, Reeves BC, Akl EA, Santesso N, Spencer FA, Shea B, Wells G, Helfand M. Non-randomized studies as a source of complementary, sequential or replacement evidence for randomized controlled trials in systematic reviews

on the effects of interventions. Res Synth Methods. 2013;4(1):49-62.
• Williamson PR, Altman DG, Bagley H, Barnes KL, Blazeby JM, Brookes ST, Clarke

제4장 문헌검색

핵심 포인트

- 체계적 문헌고찰 저자는 프로토콜 작성 시점부터 숙련된 의학사서 또는 정보 전문가와 긴밀히 협력해야 한다.
- 체계적 문헌고찰의 포함 단위는 연구보고서가 아닌 연구이지만, 대부분의 연구를 확인하고 그 연구와 그 연구결과에 대한 정보를 얻는 가장 편리한 접근방법은 연구보고서를 파악하는 것이다.
- 모든 체계적 문헌고찰시 Cochrane Central Register of Controlled Trials(CENTRAL) 및 MEDLINE 외에도 접근가능하다면 embase를 함께 검색해야 한다.
- 연구 등록부 외에도 규제 기관이나 clinical study reports(CSR)와 같은 연구의 기타 출처가 될 수 있는 정보원도 검색해야 한다.
- 검색 시 정밀도가 상대적으로 떨어져도 높은 민감도를 목표로 한다.
- 검색 전략은 너무 많은 검색 개념을 사용하지 않아야 하지만, 포함된 각 개념 내에서 다양한 검색 용어를 'OR'로 결합해야 한다.
- 프리텍스트와 MeSH 검색을 모두 사용해야 한다.
- MEDLINE에서 무작위배정 임상시험의 파악을 위한 코크란의 고감도 검색 전략(필터)을 고려해야 한다.

4.1 개요

체계적 문헌고찰을 위해서는 문헌고찰 적격기준을 충족하는 연구의 파악에 체계적이고 포괄적인 접근방식을 취해야 한다. 이 장에서는 연구 검색에 관한 몇 가지 일반적인 문제를 개략적으로 설명하고, 연구의 주요 출처를 설명하며, 검색 계획, 검색 전략 설계 및 실행, 검색 과정 중 발견된 참고문헌 관리, 검색 과정의 문서화 및 검색 결과에서 연구를 선택하는 방법을 논의한다.

이 장은 문헌고찰 저자가 검색 과정을 더 잘 이해할 수 있도록 연구 검색의 모든 측면에 대한 배경 정보를 제공하는 것을 목표로 한다. 그러나 체계적 문헌고찰 저자는 검색 과정에 대한 지원을 제공할 수 있는 숙련된 의학사서 또는 정보 전문가를 파악해야 한다. 이 장은 검색 과정에 관여하는 의학사서 및 정보 전문가에 대한 조언과 지침을 제공하는 것도 목표로 한다.

이 장은 무작위배정 임상시험들을 찾는 것에 초점을 맞춘다. 그러나 논의되는 많은 검색 원칙은 다른 연구설계에도 적용될 것이다.

4.2 일반적인 문제들

4.2.1 정보 전문가와 사서의 역할

사서와 정보 전문가는 체계적 문헌고찰 수행에 필수적인 역할을 한다. 검색 과정의 다양한 측면의 품질을 개선하기 위해 정보 전문가가 체계적 문헌고찰에 참여하는 연구가 증가하고 있다.

대부분의 체계적 문헌고찰 팀은 수행을 지원하기 위해 정보 전문가를 고용한다. 그러나 정보 전문가가 제공하는 서비스의 범위는 이용 가능한 자원에 따라 다르다. 코크란 리뷰의 경우 저자는 코크란 정보 전문가(Cochrane Information Specialist)에게 조언과 지원을 받을 수 있다. 자체 검색을 수행하는 저자는 검색해야 할 출처는 물론 실행할 정확한 전략과 관련하여 코크란 정보 전문가로부터 조언을 구해야 한다. 별도로 정보 전문가를 고용한 경우, 의학사서 또는 정보 전문가, 특히 체계적 문헌고찰을 지원한 경험이 있는 사람에게 도움을 구할 것을 권고한다.

4.2.2 비뚤림의 최소화

적절한 연구를 가능한 한 많이 파악하기 위해 광범위한 출처를 대상으로 철저하고 객관적이며 재현가능한 검색을 해야 한다. 이것은 비체계적 문헌고찰과 체계적 문헌고찰을 구별하는 주요 요소로서, 비뚤림을 최소화하고 효과와 불확실성에 대한 보다 신뢰할 수 있는 추정을 달성하는 데 도움이 된다. MEDLINE의 검색만으로는 충분하지 않다. 모든 출판된 무작위배정 임상시험을 MEDLINE에서 검색할 수 있는 것은 아니며, 관련 자료가 MEDLINE에 있더라도 이를 검색이 어려울 수 있다.

MEDLINE 외 데이터베이스에 대한 추가 검색은 가능한 한 많은 관련 연구가 확인되도록 하는 것뿐만 아니라 연구에 대한 선택 비뚤림을 최소화하기 위해서도 중요하다. MEDLINE 검색에만 의존하면 여러 출처를 더 넓게 또는 더 광범위하게 검색한 것에 비해 대표성이 떨어지는 일부 보고서만이 검색 될 수 있다.

시간과 예산의 제한으로 인해 검색의 철저함과 시간과 비용 효율성 사이에서 균형이 필요하다. 이러한 균형을 달성하는 가장 좋은 방법은 검색의 제한으로 발생할 수 있는 출판비뚤림 및 언어비뚤림을 인지하고 최소화하는 것이다. 연구의 선택이나 자료 추출과 같은 작업과는 달리, 두 사람이 동시에 독립적인 검색을 수행하는 것을 권장하지는 않는다. 그러나 모든 검색 전략은 적절한 자격을 갖춘 숙련된 의학사서 또는 정보 전문가에게 동료 검토를 받는 것은 권고한다.

4.2.3 연구 대 연구 보고서

체계적 문헌고찰은 개별 연구를 분석의 일차적 단위로 삼는다. 한 연구에서 두 개 이상의 보고서가 있을 수 있으며, 이러한 보고서 각각은 문헌고찰을 위한 유용한 정보를 제공할 수 있다. 대부분의 경우 검색 과정을 통해 검색한 개별 연구 보고서들을 대상으로 동일한 연구에 대한 다수의 보고서인지를 문헌고찰 저자가 수동으로 파악하여 이들을 서로 연관시켜야 한다. 그러나 코크란 연구 등록부나 여러 코크란 리뷰 그룹들의 특화된 등록부 및 기타 규제당국 및 산업체 기반 레지스트리 등에서 동일한 연구의 여러 보고서를 함께 제공하는 연구 기반 자료원의 수가 증가하고 있다. 연구별로 출판물을 선택하고 그룹화하기 위한 과정과 소프트웨어는 이후 섹션에서 논의한다.

4.2.4 저작권 및 라이선스

체계적 문헌고찰 저자는 저작권법 및 데이터베이스 라이선스 협약의 조건을 준수해야 한다. 연구 검색과 관련하여, 이것은 특히 데이터베이스 및 기타 출처를 검색하고 레코드를 다운로드할 때 사용 약관을 준수하고 출판물의 사본을 입수할 때 저작권법을 준수하는 것을 말한다. 저작권 법률은 국가마다 다르고 면허 계약은 조직에 따라 다르기 때문에 체계적 문헌고찰 저자는 의학사서 전문가 또는 정보 전문가로부터 이에 대한 지침을 구해야 한다.

4.3 검색원

4.3.1 서지 데이터베이스

4.3.1.1 서지 데이터베이스 소개

연구의 검색은 보고 비뚤림의 위험을 줄이고 가능한 한 최대한의 관련 근거를 파악하기 위해 최대한 광범위해야 한다(MECIR C19, C24). 일반적으로 건강 관련 서지 데이터베이스 검색은 관련 연구 분야 보고서들의 최초 집합을 파악하는 가장 효율적인 방법이다. 데이터베이스 선택은 문헌고찰 주제에 따라 이루어져야 한다. 주제가 전문적이고, 다학제적이고, 새로운 기술을 다루는 것이라면 파악을 위한 추가 데이터베이스가 필요할 수 있다.

일반적으로 임상시험 보고서를 검색하는 데 가장 중요한 자료로 간주되는 세 가지 서지 데이터베이스는 CENTRAL, MEDLINE 및 Embase이다. 코크란 리뷰에서는 CENTRAL, MEDLINE 및 Embase(체계적 문헌고찰팀이 Embase에 접근할 수 있는 경우)를 검색해야 한다.

MEDLINE 및 Embase와 같은 서지 데이터베이스는 대부분의 최신 자료에 대해 초록을 제공하고 있다. 이러한 데이터베이스의 주요 장점은 제목이나 초록의 단어에 대한 전자 검색을 제공하고, 표준화된 색인 용어 또는 통제어를 사용할 수 있다는 것이다. 코크란은 무작위 배정 임상시험 보고서의 데이터베이스를 개발하였으며 이를 Cochrane Central Register of

Controlled Trials(CENTRAL)이라는 이름으로 코크란 라이브러리에서 발행하고 있다. CENTRAL은 시작부터 코크란 리뷰에 포함시킬 수 있는 임상시험 보고서에 대한 최고의 단일 출처로 간주되어 왔다.

서지 데이터베이스는 개인이 무료로 이용하거나 필요한 만큼 지불하는 방식으로 이용가능하다. 그들은 국가차원이나 대학, 병원과 같은 기관의 사이트 라이선스를 통해서 혹은 전문가 단체의 멤버십 패키지의 일부 또는 인터넷에서 무료로 이용할 수 있다. 일부 국제기구는 데이터베이스에 대한 무료 또는 저렴한 온라인 접근 서비스를 제공한다. 세계보건기구(WHO)가 주요 출판사와 함께 설립한 HINARI(Health InterNetwork Access to Research Initiative) 프로그램은 115개 이상의 국가, 지역 및 지역의 비영리 기관에서 지역 의료 전문가를 위한 Cochrane Library를 포함한 광범위한 데이터베이스에 대한 접근을 제공한다. 또한 International Network for the Availability of Scientific Publications(INASP)는 코크란 라이브러리를 포함한 광범위한 데이터베이스 및 학술지에 대한 접근 권한을 제공한다. EIFL(Electronic Information for Libraries, EIFL)은 중부와 동부 및 남동부 유럽, 구 소련, 아프리카, 중동 및 동남아시아에 있는 60개 이상의 저소득 국가에서 학술지 및 기타 출처에 대한 합리적인 라이센스를 지원하기 위해 도서관 협약국에 기반을 둔 기구이다. 검색할 출처에 대한 추가 근거 기반 정보는 연간 두 번 갱신되는 SuRe Info 포털(http://vortal.htai.org)에서 확인할 수 있다.

4.3.1.2 MEDLINE과 Embase

체계적 문헌고찰에는 MEDLINE 검색이 포함되어야 한다. 2018년 8월 기준 MEDLINE에는 1946년 이후의 생의학 및 보건의료 관련 2500만 개 이상의 문헌이 포함되어 있다. 약 40개 국어로 된 5200개 이상의 학술지가 MEDLINE에 색인되어 있다. PubMed는 MEDLINE에 대해 아직 색인되지 않은 최신 논문도 포함하는 무료 버전의 MEDLINE이다.

MEDLINE은 또한 EBSCO, Ovid, ProQuest, STN과 같은 데이터베이스 공급업체를 통해 구독할 수 있다. 구독료를 지불하는 기관(예: 병원 및 대학)의 구성원의 경우 보통 무료로 이용가능하다. Ovid MEDLINE은 1일 차이를 두고 PubMed의 모든 이용 가능한 컨텐츠와 메타데이터를 검색할 수 있게 해준다. MEDLINE 레코드 외에도, Ovid는 PubMed에서 이용할 수 있는 모든 콘텐츠 유형을 포함한다.

Ovid 또는 PubMed 이외의 서비스 제공자 인터페이스를 통해 MEDLINE을 검색할 때에는 PubMed 대비 데이터베이스의 정확한 적용 범위에 대해 확인하는 것이 좋다.

코크란의 중재 문헌고찰에는 Embase(집필진이 Embase에 접근할 수 있는 경우) 검색도 포함되어야 한다. Embase는 2018년 6월 기준 8000여 개 학술지, 3000만 개 이상의 기록을 담고 있다. Embase는 모든 MEDLINE 레코드를 포함하므로 기술적으로 두 데이터베이스를 동시에 검색할 수 있다. Embase에는 MEDLINE에 색인되지 않은 2900개 이상의 학술지, 600만 개 이상의 레코드가 있다. Embase에는 약 90개 국 학술지에서 발행한 연구들이 포함되어 있다. Embase Classic을 통해 1947년부터 1973년까지 Excerpta Medica 인쇄 학술

지에서 디지털화된 거의 200만 개의 레코드에 대해 접근할 수 있다. Embase는 출판사 Elsevier(Embase.com)를 통해 직접 구독하거나 Ovid, ProQuest 또는 STN과 같은 다른 데이터베이스 공급업체를 경유해 구독할 수 있다. 문헌고찰 팀이 접근할 수 있는 경우 Embase를 검색하는 것이 필수적이다. Embase는 코크란이 정기적으로 검색하여 이를 CENTRAL에 반영한다.

MEDLINE과 Embase 사이의 실제 기준 중첩 정도는 주제에 따라 크게 다르지만, 두 데이터베이스의 검색을 비교하는 연구는 일반적으로 두 데이터베이스를 모두 검색해야 한다고 결론지었다.

4.3.1.3 Cochrane Central Register of Controlled Trials(CENTRAL)

서비스 개시 이후, CENTRAL은 무작위배정 임상시험 보고서의 가장 포괄적인 출처로 인정받고 있다. CENTRAL은 코크란 라이브러리의 일부로 발행되며 매월 갱신된다. 2018년 6월 현재, CENTRAL에는 127만 5천 건 이상의 임상시험 보고서 및 임상시험 레지스트리 기록을 포함하며 그 대부분은 무작위배정 임상시험이다.

CENTRAL의 레코드 중 상당수는 MEDLINE과 Embase 검색으로도 확인할 수 있다. 그러나 CENTRAL은 MEDLINE, Embase 또는 기타 서지 데이터베이스에 색인되지 않은 무작위배정 임상시험의 보고서, 디양한 언어로 출판된 참고문헌, 그리고 접근하기 어려운 학술대회 초록이나 기타 출처 등도 포함한다. 또한 임상시험 기록과 연구결과기록도 포함한다.

이러한 추가 레코드는 대부분 코크란 정보 전문가에 의해 확인되며, 이들 중 다수는 CRG의 전문 등록부를 채우기 위해 포괄적인 검색을 수행하며, 해당 분야에서 적합한 임상시험 기록을 수집한다. 이 전문 등록부는 CENTRAL에 포함되어 있다. 충분히 포괄적인 검색을 수행한 전문 등록부를 사용할 수 있는 경우, 특정 문헌고찰을 위해 CENTRAL, MEDLINE 및 Embase를 별도로 검색하는 대신 전문 등록부 검색을 수행할 수 있다. 이 경우, 검색은 더 정밀하게 이루어질 수 있지만 MEDLINE 또는 Embase와 같은 데이터베이스에 나타나는 레코드와 그 레코드들이 전문 등록부에 포함되는 사이에는 시간 지연이 있을 것이다.

CENTRAL은 코크란 라이브러리를 통해 이용할 수 있다. 많은 문헌고찰 저자는 사이트 라이센스 또는 유료 구독을 통해 접근 권한을 갖는다. 모든 코크란 정보 전문가는 CENTRAL에 접근할 수 있다.

4.3.1.4 기타 서지 데이터베이스

많은 국가와 지역은 해당 지역에서 생산된 문헌에 초점을 맞춰 다른 곳에서는 색인되지 않는 학술지와 기타 문헌을 포함하는 서지 데이터베이스를 생산한다. 또한 AMED(보완대체요법), CINAHL(간호 및 기타 보건연계 서비스) 및 PsycINFO(심리학 및 정신의학)와 같은 과목별 서지 데이터베이스도 있다. 적절한 국가, 지역 및 주제의 특정 서지 데이터베이스를 검색하는 것을 권고한다(MECIR C25).

Citation index는 표준 서지 자료 외에 특정 문헌이 참고문헌으로 인용된 경우를 기록하는

참고문헌 데이터베이스다. Citation index는 연구에 인용되는 다른 보고서에 포함될 가능성이 높기 때문에 관심 연구 보고서와 유사한 연구를 파악하는 데 사용할 수 있다.

4.3.2 진행 중인 연구 및 미발표 데이터 자료원

빠른 변화를 보이는 분야에서는 진행 중인 연구 및 미발표 데이터에 대한 접근 계획이 필요하다. 따라서 문헌고찰 저자는 의학사서 또는 정보 전문가에게 이에 대한 조언을 구해야 한다.

문헌고찰이 갱신될 때 포함 여부에 대해 평가될 수 있도록 진행 중인 연구를 파악하는 것이 중요하다. 관련성이 있을 수 있는 연구 중 현재 진행 중인 연구가 존재하는지와 완료일을 예상하면 문헌고찰의 갱신 시기에 관한 결정뿐만 아니라 문헌고찰의 완료 목표 시점에도 영향을 미칠 수 있다. 진행 중인 연구에 대한 정보는 '진행 중인 연구의 특성' 표에 포함되어야 한다.

연구가 완료되었을 때에도, 어떤 것들은 출판되지 않는다. '통계적으로 유의한' 연구결과와 출판물 사이의 연관성은 여러 연구에서 언급되었다. 미발표 연구에 대해 알아보고, 적격하고 적절한 경우 체계적 문헌고찰에 연구결과를 포함시키는 것이 비뚤림을 최소화하는 데 중요하다.

완성되었지만 출판되지 않은 연구에 대한 정보를 얻는 쉽고 신뢰할 수 있는 방법은 없다. 그러나 주요 참고 문헌 데이터베이스와 학술지 이외의 출처에서 해당 연구 및 그 연구결과에 더 잘 접근하는 몇 가지 경로가 있다. 여기에는 임상시험 등록부, 임상시험 연구결과 등록부, 규제 승인을 위해 업계에서 작성한 보고서인 규제 기관 제출용 임상 연구 보고서(CSR)가 포함된다.

미출판 데이터를 얻기 위한 방법 중에는 저자에게 이메일을 통해 연락하는 것이 다른 방법보다 더 성공적이었다. 영어 이외의 언어로 된 미출판 연구들의 포함 여부가 메타분석의 연구결과에 미치는지에 대한 평가 연구에서 해당 연구가 문헌고찰의 연구결과와 결론에 영향을 미치는 경우는 거의 없었다. 하지만 관련 연구가 거의 없는 경우에는 영향을 미칠 수 있다.

미출판 또는 진행 중인 연구에 대한 정보를 얻기 위해 관련 개인연구자와 단체에 연락하는 것을 권고한다(MECIR C31). 이는 완료되었지만 출판되지 않은 연구를 식별하는 데 사용될 수 있다. 이를 위한 한 가지 방법은 개별 연구 저자에게 관련될 수 있는 추가 연구를 알고 있는지 여부를 검토 기준과 함께 관련 문서의 포괄적인 목록을 보내 문의하는 것이다. 이러한 접근 방식은 특히 연구가 거의 없거나 활동적인 연구 그룹이 한정된 분야에서 유용할 수 있다. 다른 전문가나 제약회사나 그 분야에 관심을 가지고 다른 사람에게도 같은 편지를 보내는 것이 바람직할 수도 있다. 일부 문헌고찰 팀은 체계적 문헌고찰 프로젝트를 위한 웹사이트를 개설하고, 현재까지 확인된 연구를 열거하고, 아직 열거되지 않은 연구에 대한 정보의 제출을 요청하기도 한다.

하지만 미출판 연구에 대한 정보를 연구자들에게 문의하는 것이 항상 성과가 있는 것은 아니다. 대안으로 RIAT(Restoring Invisible and Abandoned Trials) 같은 조직은 미출판 연구의 정보에 체계적으로 접근할 수 있는 자료원 제공을 목표로 활동한다.

4.3.3 임상시험 등록부 및 임상시험 연구결과 등록부

최근 연구에 의하면 임상시험 등록부가 무작위배정 임상시험을 추가로 확인할 수 있게 도와주는 중요한 출처이다. 코크란 중재 문헌고찰은 관련 임상시험 등록부를 검색해야 한다(MECIR C27). 여러 임상시험 등록부가 있지만, ClinicalTrials.gov와 WHO 국제 임상시험 등록 플랫폼(ICTRP)이 체계적 문헌고찰을 위한 연구 파악에 가장 중요한 자원으로 간주된다. 비록 ClinicalTrials.gov 자료가 WHO ICTRP 검색 포털에 포함되어 있더라도 모든 ClinicalTrials.gov 기록이 ICTRP 검색 포털에서 성공적으로 검색되지는 않기 때문에 ICTRP를 단독으로 검색하는 것만으로는 충분하지 않다. 코크란 외에도 미국의 AHRQ와 Institute of Medicine 등 다른 기관도 임상시험 등록부를 검색할 것을 권고한다.

연구자들은 임상시험의 개시시점에 등록하고 자신들의 연구결과에 접근할 수 있게 하는 것이 중요 하다는 사실을 점점 더 많이 인식하고 있다. 그러나 관련 법률을 도입하기 시작한 일부 국가를 제외하고는 개시 혹은 이후의 단계에서 임상시험을 등록해야 하는 국제적인 법적 강제는 없다.

제약 산업을 대표하는 단체와 개별 제약 회사를 포함한 많은 조직들이 현재 진행 중인 임상시험에 대한 정보와 완료되는 임상시험 연구결과를 제공하고 있다.

4.3.4 규제 기관 자료원 및 임상 연구 보고서(CSR)

관련성이 있는 규제 기관 자료원으로는 EU 임상 시험 등록부, Drugs@FDA, OpenTrialsFDA 등이 포함된다. 임상 연구 보고서(CSR)는 약물의 시판승인을 위해 제출된 임상 시험의 방법과 연구결과에 대한 자세한 정보를 제공하는 임상 시험 보고서다. European Medicines Agency(EMA)도 CSR을 공개하고 있다.

4.3.5 기타 출처

'회색 문헌'이라는 용어는 전통적인 학술 출판물 이외의 보고서를 가리키는 데 종종 사용된다. 문헌고찰 저자는 일반적으로 학위논문 및 학술대회 초록과 같은 출처를 검색해야 한다(MECIR C28, C29, C30). 또한 인터넷 검색, 저널 수기검색 및 가능한 경우 저널의 전체 텍스트에서 전자 검색을 고려할 수 있다. 동일한 주제에 대한 기존의 문헌고찰을 검토하고 포함된 연구의 참조 목록과 관련 체계적 문헌고찰을 확인해야 한다.

4.4 검색 전략의 설계

4.4.1 검색 전략 소개

이 절에서는 검색 전략을 설계할 때 고려해야 할 몇 가지 문제를 소개한다. 검색 전략을 설계하는 것은 복잡할 수 있으며 문헌고찰 팀은 의학사서 또는 정보 전문가의 기술과 전문 지식을 활용할 수 있다. 아래에 강조된 많은 문제들은 검색의 주제 측면(예: PICO 요소)과 연구 방법(예: 무작위배정 임상시험) 모두에 관련되어 있다. 검색이 안정적이려면 두 가지 측면 모두에서 관련 기록이 누락되지 않도록 주의해야 한다. 검색을 계획할 때 고려해야 할 사항으로는 다음과 같은 것들이 있다.

- 평가 대상 중재의 특성 또는 유형
- 문헌고찰이 다루는 질문의 복잡성과 추가 개념적 틀을 고려해야 할 필요성
- 중재에 대한 평가가 이루어지는 기간(문헌고찰의 프로토콜에 명시된 기간)
- 국가 또는 지리적 특수성을 감안한 고려사항
- 문헌고찰이 무작위배정 임상시험으로 제한되는지 또는 기타 연구설계가 적용 가능한지 여부
- 검증된 방법론적 검색 필터(특정 연구설계의 경우)를 사용할 수 있는지 여부
- 미발표 데이터를 구체적으로 찾아야 하는지 여부
- 체계적 문헌고찰이 유해효과, 경제성 평가 또는 질적 연구 등의 연구설계 관련 특별한 적격 기준을 가지고 있는지의 여부. 이 경우 이러한 기준을 해결하기 위한 검색을 수행해야 한다 (MECIR C26).

근거에 기반한 검색 전략 설계 정보는 SuRe 정보 포탈(http://vortal.htai.org/)에서 확인 가능하며 매년 두 번 갱신된다.

4.4.2 검색 전략의 구조

검색 전략을 개발하는 출발점은 문헌고찰에서 검토 중인 주요 개념을 고려하는 것이며 이를 PICO로 정리할 수 있다. 포함할 연구의 적격 기준은 검색 전략에 대한 적절한 MeSH와 텍스트워드를 선택하는 데 더 도움이 될 것이다.

데이터베이스의 검색 전략 구조는 PICO와 연구설계의 적절한 요소를 포함하여 문헌고찰의 주요 개념으로 제시해야 한다(MECIR C32). 그러나 문헌고찰 임상 질문의 모든 측면을 검색하는 것은 대개 불필요하며 심지어 바람직하지 않을 수도 있다. 연구 질문이 특정 대조군 또는 중재결과를 명시할 수 있지만, 이러한 개념은 논문의 제목이나 초록에서 잘 설명되지 않을 수 있으며 종종 통제어로 잘 색인되지 않는다. 따라서 MEDLINE과 같은 일반 데이터베이스에

서 검색 전략은 일반적으로 1) 관심 건강 상태를 검색하는 용어, 즉 인구집단, 2) 평가대상 중재를 검색하는 용어, 3) 포함할 연구설계 유형을 검색하는 용어의 3가지 세트로 구성된다. 일반적으로 각 개념별로 광범위한 검색어가 수집되며, 이를 OR 불리언과 결합된 다음 각 개념에 대한 결과를 AND 불리언을 사용하여 결합하여 각 개념이 최종 검색 결과에 표시되도록 한다. 한편 복합중재, 진단검사, 공중보건 중재, 질적 연구에 대한 문헌고찰은 매우 다른 검색 접근법을 채택할 수 있다.

4.4.3 민감도 대 정밀도

체계적 문헌고찰을 위해서는 가능한 많은 관련 연구가 문헌고찰에 포함되도록 최대한 광범위한 검색을 하여야 한다. 그러나 검색 전략을 개발할 때에는 포괄성과 관련성 사이에서 균형을 잡을 필요가 있다.

검색의 속성은 흔히 '민감도'와 '정밀도'로 정량화된다(표 4-1). 민감도는 검색원에서 확인된 관련 연구보고서의 숫자를 검색원내 관련 연구보고서의 총 숫자로 나눈 값으로 정의된다. 정밀도는 확인된 관련 연구보고서의 숫자를 확인된 연구보고서의 총 숫자로 나눈 값으로 정의된다. 검색의 포괄성(또는 민감도)를 높이면 정밀도가 떨어지고 대개 관련 없는 보고서를 더 많이 검색할 수 있다.

검색은 민감도를 극대화하는 동시에 합리적인 정밀도를 위해 노력해야 한다. 데이터베이스 검색을 통해 확인된 연구보고서는 매우 신속하게 선별되어 잠재적 관련성을 확인할 수 있다. 보수적으로 추정한 분당 1-2개의 초록 읽기 속도라면 데이터베이스 검색 결과를 시간당 60-120(또는 8시간 동안 약 500-1000)의 속도로 선별할 수 있으므로, 체계적 문헌고찰에 소요되는 총 시간을 고려하면 검색과 선별에 소요되는 시간은 그다지 많지 않은 편이다.

표 4-1. 검색의 민감도 및 정밀도

	검색된 연구보고서	검색 안된 연구보고서
관련 연구보고서	관련 보고서 검색됨(a)	관련 보고서 검색 안됨(b)
관련 없는 연구보고서	관련 없는 연구보고서 검색됨(c)	관련 없는 연구보고서 검색 안됨(d)

민감도: 모든 관련 연구보고서 중 검색으로 찾아낸 관련 연구보고서의 비율(a/(a+b))
정밀도: 모든 검색된 연구보고서 중 관련있는 연구보고서의 비율(a/(a+c))

4.4.4 통제어 및 텍스트워드

MEDLINE과 Embase 그리고 다른 많은 데이터베이스는 두 개의 검색 접근방식을 조합하여 검색할 수 있다. 하나는 제목, 초록 또는 데이터베이스에서 사용할 수 있는 기타 관련 필드

에서 발생하는 용어인 텍스트워드에 기초한다. 다른 하나는 색인 전문가가 할당한 MeSH에 기초한다. 검색은 이 두 가지 접근법의 적절한 조합을 사용해야 한다(MECIR C33).

4.4.5 언어, 날짜 및 문서 형식 제한

검색은 가능한 한 많은 연구가 포함될 수 있도록 적절한 검색대상 기간과 검색원을 가지고 언어 또는 출판 상태에 의해 제한되지 않도록 보장해야 한다. 문헌고찰 저자가 검색 전략에서 출판일 및 출판형식에 제한을 사용한다면 그 이유가 타당해야 한다(MECIR C35). 예를 들어, 서신(letter)에는 초창기 시험 보고서 또는 다른 곳에서 보고되지 않은 시험에 대한 새로운 정보가 포함될 수 있기 때문에 서신을 배제하는 것이 권장되지 않는다. 또한 '논평(comments)'으로 색인된 기사 역시 추가 검토 없이 일상적으로 배제되어서는 안 된다.

보완 대체 의학 영역을 제외하고 비영어 연구를 배제해도 대부분의 체계적 문헌고찰의 결론을 바꾸지 않는다. 그러나 체계적 문헌고찰에 비영어 연구의 포함을 지지하는 연구도 있다.

또한 영어 이외의 연구를 포함하면 연구결과의 정밀도와 결과의 일반성과 적용가능성이 증가할 수 있다. 민족의 차이에 따라 의약품에 대한 치료 반응에 차이가 있을 수 있다. 또한 영어 이외의 연구를 포함하고 민감도 분석을 수행하여 긍정적 연구결과 보고에 지리적 편향이 있는지 확인할 수 있다.

언어 제한이 적용되지 않더라도 주로 영어 기록이 포함된 데이터베이스에 대한 검색으로 제한하면 관련 연구가 누락될 수 있다. 따라서 문헌고찰 저자는 출판 언어와 관계없이 가능한 모든 관련 시험 보고서를 식별하고 적격성을 평가해야 한다. 영어 이외의 연구 보고서를 번역하거나 자료 추출에 대한 도움을 필요로 하는 경우, 이를 위한 지원을 모색해야 한다.

4.4.6 부정 연구, 기타 취소 출판물, 정정기사 및 논평 확인

포함시킬 연구의 적격을 고려할 때, 일부 연구에는 오류가 포함되어 있거나 사기성이 있는 것으로 판명되었을 수 있으며, 다른 이유로 출판 이후 정정되거나 취소될 수 있다는 점을 유념하는 것이 중요하다. 문헌고찰 저자는 관련 논문 취소 기사와 정정기사에 대한 정보를 검토해야 한다(MECIR C48). 이는 문헌고찰에 포함된 '새로운' 연구와 문헌고찰에 이미 포함된 연구에(문헌고찰이 갱신될 때) 모두 적용된다. 문헌고찰 갱신의 경우, MEDLINE에서(기존에 포함된) 연구의 최신 버전을 검색하는 것이 중요하다. 정정기사(errata)는 의도하지 않은 오류를 수정하기 위해 출판된다.

4.4.7 필터 검색

검색 필터는 특정 방법론적 설계를 가진 연구와 같은 특정 유형의 연구를 검색하도록 설계

된 검색 전략이다. 인간 대상 무작위배정 임상시험을 검색할 때, 검증된 필터를 사용하여 적절한 설계로 연구를 식별해야 한다(MECIR C34). MEDLINE 및 Embase용 무작위배정 임상시험 식별 필터는 특별히 개발되었다. InterTASC Information Specialist's Subgroup Search Filter Resource (https://sites.google.com/a/york.ac.uk/issg-search-filters-resource/filters-to- identify-systematic-reviews)는 주로 방법론적 검색 필터에 초점을 맞추고 이러한 필터 중 일부에 대한 중요한 평가를 제공하는 검색 필터 컬렉션을 제공한다. 이 사이트는, 무엇보다도, 다양한 데이터베이스 및 서비스 제공 업체에서 체계적 문헌고찰, 무작위 및 비무작위 연구 및 정성적 연구를 식별하기 위한 필터를 포함한다.

4.4.8 검색 전략의 동료 검토

검색 전략의 동료 검토를 강력히 권고한다. 검색 전략의 동료 검토는 체계적 문헌고찰에 양질의 검색 전략을 설계하고 실행하는데 필요한 단계로 점차 인식되고 있다. 체계적 문헌고찰을 뒷받침하는 검색 전략에서 오류가 발생할 수 있으며 검색 전략이 항상 높은 표준에 따라 수행되거나 보고되는 것은 아니다. 검색전략의 동료 검토를 위해 근거 기반 체크리스트(PRESS Evidence-Based Checklist 같은)를 이용해서 전자 검색 전략에서 중요한 요소를 평가하는데 사용할 수 있다. 체크리스트는 검색 전략이 기술적으로 정확한지 평가할 뿐만 아니라 검색 전략이 프로토콜의 모든 관련 측면을 포괄하고 연구 질문을 적절하게 해석했는지도 점검한다. 검색 전략에 대한 동료 검토자의 이름, 자격 증명, 소속 기관을 허락을 얻어 기록해야 한다.

4.4.9 알림(Alerts)

알림은 문헌 감시 서비스, 'push' 서비스 또는 SDI(selective dissemination of information) 라고도 불린다. 이것은 특정 문헌고찰을 위한 특정 검색을 설계하고 실행하는 데 대한 보완책으로 검색 결과에 최근에 발행된 문헌이 포함되도록 하여 최신성을 유지하는 훌륭한 방법이다. 실제로 알림은 데이터베이스 플랫폼의 개인 계정에 저장된 이전에 개발된 검색 전략을 기반으로 한다(예: EBSCO의 'My EBSCOhost - Search Alerts', Ovid의 'My Search & Alerts', PubMed의 'MyNCBI - Search'). 저장된 검색전략은 데이터베이스가 새로운 정보로 갱신될 때 내용을 필터링한다. 계정 소유자는 지정된 검색 매개변수를 충족하는 새로운 게시물이 데이터베이스에 추가될 때(일반적으로 이메일을 통해) 통지된다. PubMed의 경우 알림은 매주 또는 매월 또는 실시간으로 전달되도록 설정할 수 있으며 이메일 또는 RSS 피드를 통해 이루어진다.

알림은 검색이 이루어진 이후 문헌고찰이 이루어진 분야에서 어떤 연구가 수행되는지를 모니터링할 수 있는 유용한 도구이다.

문헌을 최신 상태로 유지되도록 도와주는 또 다른 방법으로 TOC(Tranch of Contents)에

기초한 알림을 사용하는 것이 있다. 현재 관심 학술지에 게재되고 있는 내용을 감시함으로써 보다 일반적인 수준에서 최신 정보를 유지하려고 노력하는 좋은 방법이 될 수 있다. 많은 저널들, 심지어 구독으로만 이용할 수 있는 저널들도 TOC 서비스를 무료로 제공한다. 하지만 TOC만으로는 체계적 문헌고찰을 수행하는 데 필요한 다른 연구를 파악하기는 어렵고, 보완 수단 정도로 인식하는 것이 좋다.

4.4.10 검색의 타이밍

출판된 문헌고찰은 가능한 최신이어야 한다. 초기 검색일가 의도한 출판일로부터 12개월(더 바람직하게는 6개월) 이상 전이라면, 모든 관련 데이터베이스에 대한 검색을 발행 전에 다시 실행해야 한다(MECIR C37, C38). 또한 잠재적으로 적합한 연구를 식별하기 위해 검색결과를 선별해야 한다. 이상적으로는 모든 관련 연구가 문헌고찰에 완전히 통합되어야 한다. 그렇지 않은 경우, 잠재적으로 적합한 연구는 '구분을 기다리는 연구'(또는 아직 완료되지 않은 경우 '진행 연구')에 참고문헌으로 보고될 필요가 있을 것이다.

4.4.11 검색의 중지

검색 전략을 개발하는 것은 종종 반복적이고 탐구적인 과정이다. 그것은 검색어 간의 절충을 모색하고 검색의 민감도와 정밀도에 대한 전체적인 영향을 평가하는 것을 포함한다. 검색이 완료되고 검색 전략 개발이 중단될 수 있을 때를 과학적이거나 객관적인 방법으로 결정하는 것은 종종 어렵다. 언제 중단할지 결정하는 능력은 전형적으로 많은 전략을 개발하는 경험을 통해 발전한다. 새로운 기록의 검색과 관련하여 제안되는 중지의 예로는 데이터베이스 검색 전략에 몇몇 새로운 용어를 추가해도 새로운 관련 연구가 나오지 않거나 정밀도가 특정 임계치 이하로 떨어지는 경우를 생각할 수 있다. 특정 용어 또는 개념을 삭제하면 관련 연구가 누락되는 것으로 나오는 경우 중단이 적절할 수 있다. 또 다른 고려사항은 이미 도출된 근거의 양이다. 근거가 부족한 주제에서 저자들은 검색을 중단할 시점을 결정하는 데 더 신중할 필요가 있을 수 있다. 검색 개발 중단 시점을 결정하는 데 도움이 되는 많은 방법이 설명되었지만, 접근방식에 대한 공식적인 평가는 거의 없었다.

기본적인 수준에서, 전략이 적절하게 수행되고 있는지에 대한 조사가 필요하다. 한 가지 간단한 테스트는 검색을 통해 주요 논문으로 확인된 논문을 찾았는지 또는 유사한 문헌고찰에 포함된 다른 출판물을 찾았는지를 확인하는 것이다. 그러나 이러한 전략만으로는 충분하지 않을 수 있는데 그 이유는 검색 전략이 알려진 연구을 잘 식별하지만, 다른 관련 기록이 누락시킬 수 있기 때문이다. 또한 인용 검색과 참고문헌 확인을 통해 전략이 적절한지 점검할 수도 있다. 만약 그러한 추가 방법들로 이미 검색이 된 것을 다시 확인하면, 이것은 전략이 적절하게 수행되고 있을 수 있다는 하나의 신호다. 또한 Peer Review of Electronic Search

Strategies(PRESS) 근거 기반 체크리스트를 사용하여 검색 전략이 적절한지 평가해야 한다. PRESS 영역 중 일부가 적절한 설명 없이 누락되거나 우려를 불러일으킨다면 검색이 아직 완료되지 않았을 수 있다.

성능을 평가하기 위해 포획-재포획(capture-recapture) 방법 또는 상대적 회상법(relative recall technique) 같은 통계 기법을 사용할 수도 있다.

4.5 검색 과정 문서화 및 보고

문헌고찰 저자는 검색 과정이 문헌고찰에서 올바르게 보고될 수 있도록 충분히 상세하게 문서화해야 한다(MECIR C36). 모든 데이터베이스의 검색은 이것이 가능한 범위까지 재현 가능해야 한다.

4.6 연구의 선택

4.6.1 관심 단위로서의 연구(연구보고서가 아닌)

코크란 리뷰는 사전에 정해진 적격 기준을 충족하는 연구에 대한 문헌고찰이다. 한 연구는 논문, 초록, 보고서 등과 같은 다양한 형태를 취할 수 있으므로, 광범위 검색결과 잠재적으로 관련성이 있는 특정 연구에 속한 많은 연구보고서를 파악할 수 있다. 따라서 어떤 연구가 문헌고찰에 포함될 수 있는지 결정하기 위해서는 뚜렷한 두 가지 과정이 필요하다. 하나는 한 연구에 속한 여러 보고서를 함께 연결하는 것이고, 다른 하나는 다양한 보고서에서 이용할 수 있는 정보를 사용하여 어떤 연구가 포함에 적합한지 결정하는 것이다.

4.6.2 한 연구에서 여러 보고서 파악

중복 출판물은 메타분석에서 두 번 이상 포함하는 경우 상당한 비뚤림을 유발할 수 있다. 중복 출판물은 여러 형태가 있는데 동일한 연구도 있고, 다른 시점, 다른 중재결과 등을 다루는 경우도 있다. 또한 연구대상자의 수는 출판물에 따라 다를 수 있다. 중복 출판물을 검출하는 것은 어려울 수 있으며, 문헌고찰 저자의 '탐지 작업'이 필요할 수 있다.

보고서를 비교할 때 유용한 기준으로는 다음과 같은 것이 있다.

- 임상시험 식별 번호(예: ClinicalTrials.gov 식별자(NCT 번호), ISRCTN(Universal Trial Number)(ICTRP에 의해 할당됨), 스폰서로부터 받은 식별자 등
- 저자 이름(대부분의 중복 보고서에는 항상 그렇지는 않지만 공통으로 들어가는 한명 이상의 저자가 있음)
- 위치 및 세팅(특히 병원 등 기관 명칭이 지정된 경우);

- 중재의 구체적인 세부사항(예: 용량, 횟수 등)
- 연구대상자 수 및 이들의 기초자료
- 연구 날짜 및 기간(다른 표본 크기가 모집 기간 때문인지 명확히 할 수도 있음)

이러한 요인과 기타 요인을 고려한 후에도 불확실성이 남아 있는 경우, 해당 문헌의 저자에게 연락을 해야 할 수도 있다.

동일한 연구의 여러 보고서를 수집하여 각 보고서가 아닌 각 연구가 문헌고찰의 관심 단위가 되도록 해야 한다(MECIR C42). 문헌고찰 저자는 특히 두 보고서가 상반된 연구결과를 포함하는 경우 연구결과의 출처로 사용할 보고서(주 보고서)를 선택하고 정당화할 필요가 있을 것이다. 추가 결과 측정값과 연구의 설계 및 수행에 대한 귀중한 정보를 포함할 수 있으므로 다른(보조) 보고서를 폐기해서는 안 된다.

4.6.3 연구 선택을 위한 일반적인 과정

문헌고찰에 포함할 연구를 선택하는 일반적인 과정은 다음과 같으며 프로토콜에 과정이 자세히 설명되어야 한다.

1. 서지 관리 소프트웨어를 사용하여 서로 다른 소스의 검색 결과를 병합하고, 동일한 보고서, 즉 동일한 저널 제목, 볼륨 및 페이지를 보고하는 중복 레코드를 제거한다.
2. 제목과 초록을 검토하여 명백히 관련이 없는 보고서를 제거한다.
3. 관련 가능성이 있는 보고서의 전문을 검색한다.
4. 동일한 연구의 여러 보고서를 연결한다.
5. 보고서 전문을 대상으로 적합성 여부를 점검한다.
6. 적절한 경우, 연구 적격성을 명확히 하기 위해 교신저자에게 연락할 수 있다(결측 정보 또는 중재결과 결측 등 추가 정보를 동시에 요청할 수 있음). 연구가 불완전/구득불가이면, 불완전한 것으로 태그/기록되어야 하며, 문헌고찰의 '평가 대기 중인 연구' 표에 기재한다.
10. 연구 포함에 대한 최종 결정을 내리고 자료 수집을 진행한다.
11. 아직 보고되지 않은 진행 중인 모든 시험을 태그 또는 기록하여 '진행 중인 연구'표에 추가할 수 있도록 한다.

보고되지 않은 측정 중재결과 자료가 있다는 이유만으로 특정 연구를 문헌고찰에서 배제해서는 안 된다는 점에 유의한다(MECIR C40).

4.6.4 선정 절차의 이행

문헌고찰에 포함할 연구에 대한 결정은 문헌고찰 과정에서 이루어지는 가장 영향력 있는 결정 중 하나이며, 그러한 결정에는 판단이 수반된다. 각 연구가 적격 기준을 충족하는지 여부를 결정하기 위해 독립적으로 두 사람이 참여한다.

이상적으로는, 관련 없는 보고서를 삭제하기 위해 두 사람이 독립적으로 제목과 초록을 선별하는 작업을 하는 것이 바람직하나 한명이 제목과 초록을 최초로 평가하는 것은 인정된다. 단, 각 연구가 연구 보고서의 전문을 바탕으로 적격기준을 충족하는지 여부를 최종 결정하기 위해서는 독립적인 두 사람이 참여하는 것이 필수적이다(MECIR C39).

적어도 두 명의 연구자가 참여하면 관련있는 보고서가 제외되는 위험을 줄일 수 있다. 효율적인 선별을 위한 방법으로써 기계 학습적 접근법과 결합한 한명의 연구자가 선별하는 방법이 가능해질 것으로 기대된다. 적어도 한 명의 연구자가 문헌고찰 중인 분야에 정통해야 하는 것은 중요하지만, 두 번째 연구자는 내용 전문가가 아닌 사람으로 하는 것이 유리할 수 있다.

연구를 포함해야 하는지에 대한 불일치는 일반적으로 토론으로 해결할 수 있다. 종종 의견 불일치의 원인은 문헌고찰 저자 중 한 쪽의 단순한 실수인 경우가 많다. 이견차가 해석의 차이에 기인하는 경우, 제삼자에 의한 중재를 요구할 수 있다. 때때로, 추가 정보 없이는 해당 연구를 포함시킬지에 대한 이견을 해소할 수 없는 경우도 있을 것이다. 이러한 경우, 저자들은 문헌고찰에서 추가 정보를 연구 원저자로부터 얻을 때까지 평가를 기다리는 연구로 분류할 수 있다.

코크란 리뷰의 경우, 흐름도와 '배제된 연구의 특성' 표를 완성할 수 있도록 선택 과정을 충분히 상세히 문서화해야 한다.(MECIR C41). 선택 과정 동안 흐름도를 구성할 수 있도록 참고문헌 수를 추적하고 그 후에 연구의 수를 추적하는 것이 중요하다. 배제 결정과 이유는 서지관리 소프트웨어, 간단한 문서 또는 스프레드시트를 사용하거나 체계적 문헌고찰 소프트웨어를 사용하여 추적할 수 있다.

4.6.5 '배제 연구' 선택

코크란 리뷰에는 '배제 연구의 특성'이라는 배제된 연구 목록이 포함되며, 이는 포함된 연구 중에서 독자가 찾아볼 것이라 예상할 수 있는 연구의 구체적인 배제 이유를 상세히 기술하고 있다. 이는 표면적으로는 선정 기준을 충족하지만 추가 평가에서는 충족하지 못하는 모든 연구를 다룬다. 그것은 또한 모든 기준을 충족하지는 않지만 잘 알려져 있고 일부 독자들이 관련이 있다고 생각할 가능성이 있는 것들을 다룬다. 배제된 연구들을 나열하고 주요 배제의 이유를 제시함으로써, 문헌고찰의 저자들은 이러한 연구에 대한 고려를 하였음을 보여줄 수 있다. 배제된 연구의 목록 문헌고찰의 선정 기준을 명백히 충족하지 못하는 연구에 대해 '연구의 종류', '연구대상자의 종류', '중재의 종류'과 같은 항목을 제시할 필요는 없다. 광범위한 검색으로 확인된 모든 보고서를 나열해서는 안 된다. 특히, 문헌고찰이 무작위배정 임상시험

만을 포함하는 경우, 명백한 비무작위 연구를 나열해서는 안 된다. 무직위비교 임상시험을 포함한 코크란 라이브러리의 중재 문헌고찰 중 60%의 샘플에서 조사한 바에 따르면 '배제 연구의 특성' 표에 나열된 평균 연구 수는 30개이다.

4.6.6 연구 선택을 위한 소프트웨어 지원

체계적 문헌고찰에서 적격 연구를 광범위하게 검색하면 수작업으로 선별해야 하는 수천 개의 기록이 검색된다. 이러한 기록 내에서 연구를 선택하는 것은 체계적 문헌고찰을 수행하는 데 특히 시간이 많이 걸리고, 힘들고, 논리적으로 어려운 측면이 될 수 있다. 이러한 문제들을 지원하기 위해 다양한 소프트웨어 도구와 패키지가 개발되었다.

전반적으로 연구 선택을 지원하는 소프트웨어는 다음과 같이 분류할 수 있다.

- 연구 선택 과정을 지원하는 시스템(일반적으로 복수의 문헌고찰 저자가 참여하는 시스템).
- 텍스트 마이닝 및/또는 기계 학습을 기반으로 하는 도구 및 기법. 이는 선택 과정을 반자동 또는 완전 자동화하는 것을 목표로 한다.

텍스트 마이닝 도구를 포함한 체계적 문헌고찰의 다른 단계와 함께 선택 과정을 지원하는 소프트웨어는 체계적 문헌고찰 도구 상자(Systematic Review Toolbox)를 사용하여 확인할 수 있다. SR Toolbox는 체계적 문헌고찰을 지원하는 도구의 커뮤니티 기반 웹 기반 카탈로그이다.

4.6.6.1 선택 과정 관리 소프트웨어

복수의 문헌고찰 저자가 참여하는 대규모의 체계적 문헌고찰에서는 선택 과정을 관리하는 것이 어려울 수 있다. 기본적인 도구로 워드 프로세서, 스프레드시트, 서지관리 소프트웨어 등이 도움이 될 수 있으며, 이외에 연구 선택 과정을 지원하는 몇 가지 특수 제작된 시스템도 사용할 수 있다. 연구 선택을 지원하는 도구에는 다음과 것이 있다.

- Abstrackr - 기계 학습 기법을 사용하여 레코드 선별을 수행할 수 있는 무료 웹 기반 선별 도구.
- Covidence - 체계적 문헌고찰을 수행하기 위한 웹 기반 소프트웨어 플랫폼으로, 제목 및 초록 선별의 협업, 전문 검토, 비뚤림위험 평가 및 자료 추출 지원 등의 기능이 있다. 이 시스템 전체 기능을 이용하려면 유료 구독이 필요하지만 코크란 리뷰 저자에게는 무료다. 코크란 리뷰 저자가 아닌 사람을 위한 무료 평가판도 이용할 수 있다.
- DistillerSR - 서지 레코드 선별 및 자료 추출 작업을 위한 웹 기반 소프트웨어로 진행 상황의 추적기능, 평가자 간 신뢰도 평가기능, 자료를 내보낼 수 있는 다양한 관리 기능을 갖추고 있다. Cochrane과 Cambell 리뷰시 할인가능.
- EPPI-Reviewer -서지 관리, 선별, 비뚤림 평가 위험, 자료 추출 및 합성을 포함한 체계적 문헌고찰 과정의 모든 단계를 지원하도록 설계된 웹 기반 소프트웨어이다. 이 시스템은

Cochrane과 Campbell 리뷰에 무료로 사용되며, 그렇지 않으면 유료 구독이 필요하다. 무료 시험판 사용이 가능하다.

- Rayyan - 공동 선별 심사 및 전문 선택을 위한 웹 기반 애플리케이션이다. 이 시스템은 현재 무료로 이용할 수 있다.

문헌고찰 과정에서 사용되는 다른 소프트웨어 도구(예: RevMan)와의 호환성 여부는 도구를 선택할 때 고려할 수 있다. Covidence와 EPPI-Reviewer는 Cochrane이 선호하는 도구이며 RevMan과의 통합이 될 가능성이 있다.

4.6.6.2 선택 과정의 자동화

기계학습과 텍스트 마이닝을 통한 연구 선택 과정 자동화에 대한 연구는 최근 몇 년간 상당한 관심을 받아 문헌고찰 연구자가 고려해야 할 다양한 도구와 기법을 개발하게 되었다. 자동화된 도구를 사용하면 연구 선택과 관련된 작업 부하를 크게 줄일 수 있다. 예를 들어, 연구에 따르면 자동화를 채택하면 민감도를 최대 5%까지 감소시키지만 수동 선별의 필요성이 최소 30%, 어쩌면 90% 이상 감소할 수 있다고 한다.

대규모 데이터 세트에서 학습하는 것 외에도, 기계 학습 모델에서는 개별적 문헌고찰에 적격 기준을 적용하는 방법을 배울 수 있다. 이 접근방식은 '능동 학습'이라는 과정을 사용하며, 결과 목록의 상단에 관련될 가능성이 가장 높은 기록을 지속적으로 홍보하여 연구 선택을 반자동화할 수 있다. 다음의 여러 소프트웨어 도구가 '능동적 학습'을 지원한다.

- Abstrackr(http://abstrackr.cebm.brown.edu/)
- Colandr(https://www.colandrapp.com/)
- EPPI-Reviewer(http://eppi.ioe.ac.uk/)
- Rayyan(http://rayyan.qcri.org/)
- RobotAnalyst(http://nactem.ac.uk/robotanalyst/)
- Swift-review(http://swift.sciome.com/swift-review/)

4.7 문헌 검색 관련 MECIR

C19. 검색의 기획(필수)	
연구 확인에 사용할 방법을 미리 계획한다. 적격기준에 부합하는 가능한 한 많은 연구를 찾아낼 검색을 설계하여 적절한 검색연도와 출처를 대상으로 언어 또는 출판 상태에 의해 제한되지 않도록 보장한다.	검색은 문헌고찰의 선정 기준에 맞게 이루어져야 하며, 검색을 계획할 때에는 모든 유형의 적절한 연구를 고려하는 것이 중요하다. 검색이 출판 상태 또는 출판 언어로 제한되는 경우, 출판 비뚤림 또는 언어 비뚤림의 가능성이 있다. 영어 데이터베이스에서 언어 제한을 없애는 것은 비영어 학술지와 데이터베이스를 검색하는 데 대한 좋은 대안이 되지 못한다.
C24. 일반적인 서지 데이터베이스 및 CENTRAL 검색(필수)	
CRG(Cochrane Review Group)의 전문 등록부(예를 들어 Cochrane Register of Study 또는 CENTRAL)를 검색한다.(CRG 또는 문헌고찰 저자가 Embase를 사용할 수 있는 경우) CENTRAL, MEDLINE 및 Embase를 검색했는지 확인한다.	연구에 대한 검색은 출판 비뚤림의 위험을 줄이고 가능한 한 관련된 근거를 파악하기 위해 최대한 광범위해야 한다. 포함할 최소 데이터베이스는 CRG의 전문 등록부, CENTRAL, MEDLINE 및 Embase이다. 불필요한 노력의 중복을 피하기 위해 전문가가 필요할 수 있다. MEDLINE, Embase 및 CRGs의 전문 등록부의 적격 연구에 대한 일부 보고서는 이미 CENTRAL에 포함되어 있다.
C25: 특정 전문분야 서지 데이터베이스 검색수행(권고)	
적절한 국가, 지역 및 주제별 참고 문헌 데이터베이스를 검색한다.	연구에 대한 검색은 출판 비뚤림의 위험을 줄이고 가능한 한 관련 근거를 파악하기 위해 최대한 광범위해야 한다. 문헌고찰 주제와 관련된 데이터베이스(예: 간호 관련 주제에 대한 CINAHL, 심리 개입에 대한 PsycINFO)를 다루어야 하며, 지역 데이터베이스(예: LILACS)를 고려해야 한다.
C26: 다른 종류의 근거 검색(필수)	
고찰이 유해효과, 경제성 평가 또는 질적 연구 등의 연구설계와 관련하여 구체적인 적격기준을 가지고 있는 경우 이를 해결하기 위한 검색을 수행한다.	때로는 유해효과 관련 질문을 평가하기 위해 비무작위 연구 또는 경제성 평가 연구와 같은 다양한 유형의 근거에 대해 서로 다른 검색을 수행할 수 있다.
C27: 임상시험 등록부를 검색하기(필수)	
ClinicalTrials.gov, WHO ICTRP 포털 및 기타 자료원 등 주제와 관련된 임상시험의 등록부를 검색한다.	연구에 대한 검색은 출판 비뚤림의 위험을 줄이고 가능한 한 관련 근거를 식별하기 위해 최대한 광범위해야 한다. ClinicalTrials.gov는 WHO ICTRP 포털 내의 등록부 중 하나로 포함되지만 ClinicalTrials.gov 만의 추가 기능 때문에 ClinicalTrials.gov와 ICTRP 포털 모두 별도로 검색하는 것이 좋다.
C28: 회색문헌 검색(권고)	
보고서, 학위논문, 학술대회 초록과 같은 관련 회색 문헌 출처를 검색한다.	연구의 검색은 출판 비뚤림의 위험을 줄이고 가능한 많은 관련 근거를 파악하기 위해 가능한 광범위하게 이루어져야 한다.

C29: 다른 문헌고찰 내용 검색(권고)	
동일 주제에 대한 이전 문헌고찰의 내용을 검색	연구의 검색은 출판 비뚤림의 위험을 줄이고 가능한 많은 관련 근거를 파악하기 위해 가능한 광범위하게 이루어져야 한다.
C30: 참고문헌 목록 검색(필수)	
포함된 연구의 참고문헌 목록과 파악된 관련 체계적 문헌고찰을 확인	연구의 검색은 출판 비뚤림의 위험을 줄이고 가능한 많은 관련 근거를 파악하기 위해 가능한 광범위하게 이루어져야 한다.
C31: 관련 개인연구자나 단체와 접촉하여 검색하기(권고)	
출판되지 않았거나 진행 중인 연구에 대한 자세한 내용을 관련 개인 및 단체에 문의한다.	연구에 대한 검색은 출판 비뚤림의 위험을 줄이고 가능한 한 관련 근거를 파악하기 위해 최대한 광범위해야 한다. 문헌고찰이 갱신될 포함될 수 있도록 현재 진행 중인 연구를 파악하는 것이 중요하다.
C32: 서지 데이터베이스의 구조화 검색 전략(필수)	
PICO 및 연구설계의 적절한 요소를 사용하여 문헌고찰의 주요 개념을 중심으로 서지 데이터베이스의 검색 전략을 만든다. 검색을 구조화할 때 민감도를 극대화하는 동시에 합리적인 정밀도를 위해 노력해야 하며 'AND' 및 'OR' 불리언을 올바르게 사용하는지 확인한다.	부적절하거나 불충분한 검색 전략은 서지 데이터베이스에 포함된 문헌을 확인하지 못할 수 있다. CRG의 정보 전문가로부터 전문 지식을 구할 필요가 있을 수 있다. 검색 전략의 구조는 문헌고찰에서 검토하는 주요 개념에 기초해야 한다. MEDLINE과 같은 일반적인 데이터베이스에서 연구를 파악하기 위한 검색 전략은 일반적으로 1) 관심 건강 상태 검색 용어, 즉 인구집단, 2) 평가된 중재 검색 용어, 3) 포함할 연구설계 유형을 검색하는 용어 세 가지 용어를 포함할 것이다. 그러나 예외도 있다. 예를 들어, 복합중재에 대한 문헌고찰일 경우, 인구집단 또는 개입만을 검색해야 할 수 있다. 각 개념 내에서 용어는 'OR' 불리언과 결합되며, 개념들은 'AND' 불리언과 결합된다. 'NOT' 불리언은 검색 세트에서 관련 기록을 실수로 제거할 위험이 있어 가능한 한 피해야 한다.
C33: 서지 데이터베이스에서 검색전략을 개발하기(필수)	
적절한 통제어(예: 'exploded' 용어를 포함한 MESH, Emtree) 및 프리텍스트 용어(예: 철자 변형, 동의어, 약어, 절단어 및 근접 불리언을 고려함)를 파악한다.	부적절하거나 불충분한 검색 전략은 서지 데이터베이스에 포함된 문헌을 파악하지 못할 수 있다. 검색 전략은 각 데이터베이스에 맞춰 수정되어야 한다. MeSH 용어는 관련 연구를 놓치지 않기 위해 적절한 곳에 'exploded'되는 것이 중요하다. Embase 검색시 Emtree와 다른 여러 데이터베이스에도 동일한 원리가 적용된다. MEDLINE과 Embase에 대한 통제어의 검색이나 색인은 동일하지 않다. 가능한 한 포괄적으로 하기 위해서는 선택한 각 개념에 대해 폭넓은 프리텍스트 용어를 포함시킬 필요가 있다. 여기에는 자르기 및 와일드카드의 사용이 포함될 수 있다. 검색전략을 개발하는 것은 이미 검색된 것을 기초로 하여 사용된 용어를 수정하는 반복적인 과정이다.

C34: 검색 필터 사용(권고)	
MEDLINE에서 무작위배정 임상시험의 파악을 위한 Cochrane 고감도 검색 전략을 포함하여 특별히 설계되고 테스트된 검색 필터를 사용하되, 사전 필터링된 데이터베이스에는 필터를 사용하지 않는다(예: CENTRAL에 무작위배정 임상시험 필터 사용 또는 DARE에 체계적 문헌고찰 필터 사용).	부적절하거나 불충분한 검색 전략은 서지 데이터베이스에 포함된 기록을 파악하지 못할 수 있다. 검색 필터는 주의해서 사용해야 한다. 개발 및 보고된 성능의 신뢰도뿐만 아니라 데이터베이스에 영향을 미치는 인터페이스와 색인 변경이 빈번하게 발생하는 점을 감안할 때 현재의 정확도, 관련성 및 효과에 대해 평가해야 한다.
C35: 데이터베이스 검색 제한(필수)	
출판일 및 출판형식에 대한 제한적 검색 전략을 정당화한다.	검색의 출판일 제한은 연구의 적격 기준에 날짜 제한이 있을 때만 사용해야 한다. 예를 들어, 특정 시점 이후에만 개입이 가능한 경우, 관련 연구가 특정 기간 동안에만 보고될 수 있었다는 사실이 알려진 경우에만 적용되어야 한다. 문헌고찰에 대한 갱신 검색은 중복 작업을 피하기 위해 데이터베이스에(출판일보다는) 진입하는 날짜로 제한될 수 있다. 적합한 연구로서 가치가 있을 수 있으므로 출판 형식 제한(예: 기고문 배제)은 일반적으로 코크란 리뷰에서 사용해서는 안 된다.
C36: 검색 과정 문서화(필수)	
검색 과정을 충분히 상세히 문서화하여 문헌고찰시 정확하게 보고할 수 있도록 한다.	검색 과정(검색된 소스, 검색 연구대상자, 사용 대상 용어 포함)는 모든 데이터베이스의 검색이 재현 가능한 범위까지 문헌고찰에서 정확하게 보고될 수 있도록 과정 전체에 걸쳐 충분히 상세하게 문서화해야 한다.
C37: 검색의 재실행(필수)	
문헌고찰 또는 문헌고찰 갱신을 게시하기 12개월 전에 모든 관련 데이터베이스를 검색하거나 갱신하고 잠재적으로 적합한 연구가 있는지 결과를 확인한다.	출판된 문헌고찰은 최대한 최신근거를 포함해야 한다. 초기 검색일가 예정된 출판일로부터 12개월(바람직하게는 6개월) 이상이고, 검색결과 잠재적으로 적합한 연구가 선별되었다면 이상적으로는 해당 연구가 문헌고찰에 완전히 통합되어야 한다. 그렇지 않은 경우, 잠재적으로 적합한 연구는 최소한 '구분을 기다리는 연구'로 보고해야 한다.
C38: 재실행 검색 결과 통합(권고)	
문헌고찰 또는 문헌고찰 갱신이 발표되기 12개월 전에 검색을 재실행하고 확인된 모든 연구를 문헌고찰에 통합한다.	발표된 문헌고찰은 가능한 최신이어야 한다. 검색을 재실행한 후, 문헌고찰에 완전히 새로운 연구를 통합할 것인지의 결정은 출판의 지연과 균형을 맞출 필요가 있다.
C39: 포함 결정(필수)	
각 연구가 적격 기준을 충족하는지 여부를 결정하기 위해 독립적으로 작업하는 두 사람이 참여하고 의견 불일치를 해결하기 위한 과정을 미리 정의한다.	연구 선택 과정을 중복해서 운영하면 실수할 위험과 선택이 한 사람의 편견에 의해 영향을 받을 가능성을 모두 줄일 수 있다. 포함 결정은 가능한 경우, 대개 제목과 초록을 바탕으로 최초 선별을 한 후 전체 본문에 기초해 시행한다. 두 사람이 독립적으로 이 최초 심사를 수행하는 것은 바람직하지만 의무적이지는 않다.

C40: 사용 가능한 자료가 없는 연구 배제(필수)	
측정된 중재결과 자료가 '사용 가능한' 방식으로 보고되는지 여부에 관계없이 연구를 문헌고찰에 포함시킨다.	체계적 문헌고찰은 일반적으로 관련 중재의 적격 연구설계에 포함되었고 측정된 관심 중재결과가 있는 모든 관련 연구대상자를 포함하도록 모색해야 한다. 선별적 중재결과 보고 및 체계적 문헌고찰 과정을 훼손할 위험이 있기 때문에 문헌고찰에서 중재결과 자료의 보고에만 근거하여 연구를 배제해서는 안 된다. 그러한 연구는 메타분석에 포함될 수 없지만 누락의 영향을 고려해야 한다. 중재결과가 측정되지 않은 연구는 합법적으로 배제될 수 있다.
C41: 파악된 연구에 대한 결정 문서화(필수)	
흐름도 및 '배제 연구의 특성' 표를 완성할 수 있도록 선택 과정을 충분히 상세히 문서화한다.	검색으로 파악된 모든 연구에 대한 결정을 문서화해야 한다. 많은 수의 연구는 단지 제목과 초록만으로 배제하는 데 충분하다. 초기 선별에서는 잠재적으로 적격으로 판단되는 연구에 대해 대략적인 분류를 하는 것만으로도 충분하다. '배제 연구의 특성' 표에 나열된 연구는 사용자가 문헌고찰 중 연구에 포함될 수 있을 것이라 합리적으로 기대되는 연구들이며 따라서 이들을 배제하는 명확한 이유를 하나 이상 문서화해야 한다. 저자는 각 문헌고찰에 대해 보고서를 연구에 연계할 시기를 결정해야 한다(복수의 보고서가 하나의 연구에 속하는 경우). 포함 및 배제된 연구 목록은 보고서보다는 연구에 기초해야 한다.
C42: 여러 보고서 모으기(필수)	
각 보고서가 아닌 각 연구가 문헌고찰의 관심 단위가 되도록 동일한 연구의 여러 보고서를 수집한다.	동일한 연구의 복수 보고서를 복수의 연구인 것처럼 생각하는 것은 잘못된 것이다. 그러나 연구의 보조 보고서는 설계와 수행에 대한 귀중한 정보를 포함할 수 있으므로 폐기해서는 안 된다. 문헌고찰 저자는 연구결과의 출처로 사용할 보고서를 선택하고 정당화해야 한다.
C48: 정정기사 검사(필수)	
관련된 모든 논문 취소 문구와 정정기사를 검사하여 정보를 확인한다.	일부 연구는 부정으로 밝혀졌거나 다른 이유로 출판된 이후 논문 취소되었을 수 있다. 정정기사는 포함된 연구에서 중요한 한계 또는 심지어 치명적인 결함을 드러낼 수 있다. 이러한 모든 것은 문헌고찰 또는 메타분석에서 연구의 배제로 이어질 수 있다. 인용 자료와 함께 적절한 필드를 다운로드하여 모든 데이터베이스 검색에서 이 정보가 검색되도록 주의해야 한다.

4.8 참고문헌과 읽을거리

- Lefebvre C, Glanville J, Briscoe S, Littlewood A, Marshall C, Metzendorf M-I, Noel-Storr A, Rader T, Shokraneh F, Thomas J, Wieland LS. Chapter 4: Searching for and selecting studies. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA(editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.0(updated July 2019). Cochrane, 2019. Available from www.training.cochrane.org/handbook.
- ClinicalTrials.gov(https://clinicaltrials.gov/ct2/home)
- Cochrane Central Register of Controlled Trials(https://www.cochranelibrary.com/central/about-central)
- Cochrane Information Specialists(https://training.cochrane.org /resource/cochrane-information- specialists-handbook)
- InterTASC Information Specialist's Subgroup Search Filter Resource(https://sites.google.com/a/york.ac.uk/issg-search-filters-resource/filters-to-identify-systematic-reviews)
- Restoring Invisible and Abandoned Trials(https://restoringtrials.org/whatisriat)
- SuRe info 포털(http://vortal.htai.org)
- WHO International Clinical Trials Registry Platform(https://www.who.int/ictrp/en)

제5장 자료 수집

핵심 포인트

- 임상시험 등록부, 규제문서, 임상 연구 보고서와 같은 자료원을 더 많이 이용할 수 있게 됨에 따라 동일 연구에 대한 여러 보고서를 확인하여 자료 추출 전후에 연계할 필요가 있다.
- 문헌고찰 저자는 자료수집 양식을 용이하게 설계하기 위해 문헌고찰에서 제시할 표와 그림의 개요를 개발하는 것이 좋다. 성공적인 자료수집의 핵심은 사용하기 쉬운 양식을 구성하고 모호하지 않은 자료를 수집하는 것이다.
- 자료는 계산되거나 변환될 수 있으며 메타분석에 필요한 자료가 무엇인지 확인하는 노력도 필요하다.
- 자료는 향후 접근과 공유가 가능한 형태로 수집하고 보관해야 한다.

5.1 서론

체계적 문헌고찰의 목적은 연구 질문과 관련된 모든 연구를 확인하고 설계, 비뚤림위험, 연구결과에 대한 자료를 종합하는 것이다. 체계적 문헌고찰을 위해 수집된 자료는 정확하고 완전해야 하며 향후 문헌고찰 업데이트 및 자료 공유를 위해 접근할 수 있어야 한다. 이 장에서는 학술지 논문과 기타 연구 보고서에서 직접 자료를 추출하는 것을 포함하여 자료 수집을 위한 접근법을 설명한다.

5.2 자료 출처

연구는 다양한 출처에 보고된다. 동일한 연구에 대한 여러 보고서를 함께 연결하는 것이 중요하다.

- **학술지 논문**은 체계적 문헌고찰에 포함된 자료 대부분의 출처이다. 하나의 연구는 연구의 일부 측면(예 : 설계, 주요 연구결과 및 기타 연구결과)에 더 중점을 주면서 여러 학술지 논문으로 보고 될 수 있다.
- **학술대회 초록**은 일반적으로 이용 가능하나 학술대회 초록에 제시된 정보는 신뢰도, 정확성, 상세 정도 등의 수준이 매우 다양하다.
- **정정기사(Errata)와 서신**은 논문의 중요한 약점 및 논문 취소를 포함하여 연구에 관한 중요한

자료원이 될 수 있기 때문에 이 부분이 확인되면 검토해야 한다.

- **임상시험 등록부(Trials registers)**(예: ClinicalTrials.gov)은 계획 중이거나 시작된 임상시험 목록(catalogue)이다. 임상시험을 확인하고 발표된 중재결과와 연구 계획을 비교하기 위한 중요한 출처가 된다.
- **임상 연구 보고서(Clinical study reports, CSRs)**에는 ICH(International Conference on Harmonisation)에서 규정한 구성과 내용 지침에 따라 임상 문제, 설계, 수행, 임상시험의 연구결과에 대한 포괄적인 설명이 포함되어 있다. 임상 연구 보고서는 프로토콜, 통계 분석 계획, 표본 증례 보고서 양식 및 환자 자료 목록(모든 심각한 이상 반응에 대한 기술 포함) 등이 포함되어 있는 부록뿐만 아니라 표와 그림도 포함되어 있어 길이가 수천 페이지가 될 수 있다. 임상 연구 보고서에는 다른 자료 출처보다 임상시험 방법 및 결과에 대한 더 많은 자료가 포함된다. 임상 연구 보고서는 접근하기 어렵고 보통 공개적으로 사용할 수 없다. 문헌고찰 저자는 유럽 의약품청(European Medicines Agency)에 임상 연구 보고서를 요청할 수 있다. 미국 식약청(Food and Drug and Administration)은 2018년 시범 프로그램을 시작하여 신약 신청을 위한 임상 연구 보고서의 일부를 기관 웹 사이트에 게시했다. 많은 임상 연구 보고서는 봉인되지 않은 소송 문서, 저장소(예 : clinicalstudydatarequest.com), 기타 공개 자료, 자료 공유 채널(예. 예일 대학교 Open Data Access Project) 등을 통해서도 얻을 수 있다.
- **규제 심사자료(Regulatory reviews)**는 미국 식약청(Food and Drug Administration)이나 유럽 의약품청(European Medicines Agency)에서 제공하는 것과 같이 시장 승인을 위해 제조업체가 제출한 의약품, 생물 제제 및 의료 기기에 대한 유용한 정보를 제공한다. 이 문서는 최초 임상시험 자료를 재분석 한 후 마케팅을 위해 제품을 승인하는 과정에서 회사 직원이 준비한 임상 연구 보고서 및 관련 문서의 요약이다. 미국 식약청의 규제 심사자료를 예로 들면, 1997년부터 승인된 의약품에 대한 의약품 승인 패키지를 해당 기관 웹 사이트에서 이용할 수 있다. 1997년 이전에 승인된 약에 대해서는 정보는 요청해야 한다. 약물 승인 패키지에는 승인서, 의학 심사자료, 화학 심사자료, 임상 약리 심사자료 및 통계 심사자료와 같은 다양한 문서가 포함되어 있다.
- **개별 연구대상자 자료(Individual participant data, IPD)**는 일반적으로 연구 책임자로부터 직접 얻거나 공개 자료 저장소(예 : www.clinicalstudydatarequest.com)에서 확인할 수 있다. 이 자료에는 일반적으로 각 연구대상자의 특성, 중재(또는 노출)군, 예후 요인(prognostic factors) 및 중재결과 측정치 등의 변수가 포함되어 있다. IPD에 대한 접근을 통해 문헌고찰 저자가 프로토콜에 설명된 선호 분석 방법에 따라 유연하게 자료를 재분석할 수 있고, 문헌고찰에 포함된 여러 연구들의 다양한 분석의 문제점을 해소할 수도 있다.

5.2.1 관심 단위로서 연구

체계적 문헌고찰에서 보고서보다는 연구가 주요 관심 단위이다. 연구가 여러 출처에서 보고되었을 수 있으므로 포괄적으로 검색하면 많은 보고서를 확인할 수 있다. 반대로 한 보고서에 다수의 연구를 기술 할 수 있다. 동일한 연구에 대한 여러 보고서를 함께 연결해야 한다. 일부

저자는 자료수집 전에 보고서를 연결하고 보고서 전체에서 단일 양식으로 자료를 수집하는 것을 선호한다. 다른 저자는 각 보고서에서 자료를 수집한 다음 수집한 자료를 통해 연결하는 것을 선호한다. 보고서의 특성에 따라 두 전략 중 어느 하나가 더 적절할 수 있다. 자료수집이 시작될 때까지 두 연구 보고서가 동일한 연구와 관련 있다는 것이 명확하지 않을 수도 있다.

같은 연구의 여러 보고서를 연결하는 것이 어려울 수 있으며 체계적 문헌고찰 저자는 '탐색 작업(detective work)'을 해야 할 수도 있다. 동일한 임상시험에 대한 여러 출처가 서로를 참조하지 않거나 저자를 공유하지 않거나 연구설계, 특성, 중재결과, 연구결과에 관한 정보가 불일치 할 수도 있다.

보고서를 연결하는 데 가장 유용한 기준은 다음과 같다.

- 임상시험 등록 번호
- 저자 이름
- 연구 스폰서 및 스폰서 식별자(예 : 보조금(grant) 또는 계약 번호)
- 위치 및 세팅(특히 병원과 같은 기관 이름이 지정된 경우)
- 중재의 구체적인 세부 사항(예 : 복용량, 빈도)
- 연구대상자 수 및 기초자료
- 연구 날짜와 기간(다른 표본 크기가 모집 기간이 달라서 생긴 것인지 여부를 명확히 할 수 있음), 추적 기간, 하위군

문헌고찰 저자는 가능한 한 많은 임상시험 특성을 사용하여 여러 보고서를 연결해야 한다. 이러한 요소를 고려한 후에도 불확실성이 남아 있다면, 확인을 위해 연구 저자 또는 후원자와 연락할 필요가 있다.

5.2.2 출처 결정하기

모든 가능한 출처에서 모든 적합한 연구를 확인하기 위한 포괄적인 검색은 자원 집약적이지만 고품질의 체계적 문헌고찰에 필요하다. 일부 자료 출처는 다른 자료 출처보다 유용하기 때문에 체계적 문헌고찰 저자는 사용 가능한 자료 출처와 문헌고찰에 가장 유용한 정보가 포함될 수 있는 자료 출처를 고려해야 한다. 이러한 고려 사항은 프로토콜에 설명되어야 한다.

임상 연구 보고서 및 개별 연구대상자 자료(IPD)에 대한 접근 권한을 얻는 데 종종 시간이 오래 걸릴 수 있다.

5.2.3 연구자와 교신

체계적 문헌고찰 저자는 연구설계의 세부 사항, 측정된 연구결과 전체 및 수치 결과에 대해

모든 정보를 얻을 수 없는 경우가 종종 있다. 그러한 경우, 저자는 연구자에게 연락하는 것이 좋다. 연구 보고서에서 얻을 수 없는 경우 연구자의 연락처는 대학이나 기관 직원 목록, 전문적인 사회집단의 회원 명단 또는 웹 검색으로 얻을 수 있다. 연구 보고서에 이름이 기재된 담당자에게 연락을 취하거나 응답하지 않는 경우 다른 저자에게 연락하는 것이 좋다.

체계적 문헌고찰 저자는 필요한 정보의 성격을 고려하여 적절하게 요청해야 한다. 임상시험 진행에 대한 설명 자료를 얻으려면 개방형 질문(예 : 배정 과정 수행 방법 또는 결측치 처리 방법)으로 요청하는 것이 적절할 수 있다. 특정 수치 자료가 필요한 경우 구체적으로 요청하고 짧은 자료수집 양식을 제공하는 것이 도움이 될 수 있다. IPD가 필요한 경우 구체적으로 요청해야 한다.

5.3 수집 자료

5.3.1 자료의 정의

'자료'란 연구방법, 연구대상자, 세팅, 맥락(context), 중재, 중재결과, 연구결과, 출판 및 연구자에 대한 세부 정보를 포함한 연구에 관한 정보(또는 파생된 정보)로 정의할 수 있다. 체계적 문헌고찰을 위해 어떤 자료가 필요할지 사전에 계획하고 그것을 얻기 위한 전략을 수립해야 한다. 소비자 및 기타 이해 관계자의 참여는 수집된 자료의 범주가 체계적 문헌고찰 사용자의 요구와 충분히 부합하는지 확인하는 데 도움이 될 수 있다. 조사할 자료는 프로토콜에 기술해야 한다.

수집된 자료를 통해 포함된 연구를 적절히 기술하고 표와 그림의 구성을 작성하며, ROB 평가, 합성 및 메타분석이 가능해야 한다. 문헌고찰 저자는 체계적 문헌고찰을 위한 보고지침에 익숙해져야 한다. 체계적 문헌고찰의 정보 유형은 아래 표 5-1에 요약되어 있다.

표 5-1. 체계적 문헌고찰의 정보 유형

- **자료 추출에 대한 정보:** 자료 추출자 이름, 자료 추출일, 각 보고서의 특징 확인
- **적격기준:** 문헌고찰을 위한 연구의 적격성 확인, 배제 사유
- **연구 방법**
  - 무작위배정 임상시험인 경우 평행(Parallel), 요인, 교차, 군집설계 중 무엇인지, 비무작위 연구인 경우 연구설계 특성
  - 단일 또는 다기관 연구; 만일 다기관인 경우 모집 센터 수
  - 사용된 모집 절차, 표본추출 절차(개별 연구대상자 및 군집/장소의 수준 포함)
  - 등록 시작과 종료일; 연구대상자 추적 기간
  - 무작위배정 임상시험인 경우 무작위 순서생성, 배정순서 은폐, 눈가림에 대한 상세내용, 비무작위 연구인 경우 교란, 선택 비뚤림, 정보 비뚤림을 방지하고 제어하는 데 사용된 방법
  - 결측 자료를 예방하고 처리하는 데 사용된 방법
  - 통계 분석 : 분석 단위(예 : 개별 연구대상자, 클리닉, 마을, 신체 일부), 사용된 통계 방법
  - 보고 및 기타 비뚤림의 가능성
  - 연구 자금 지원 또는 기타 물질적 지원의 출처
  - 저자의 재정적인, 비재정적인 이해 상충
- **연구대상자**
  - 세팅
  - 연구대상자의 지역 및 국가
  - 연구대상자의 적격기준
  - 연구 시작 시점에서 연구대상자의 특성(예 : 연령, 성별, 동반 질환, 사회경제적 상태)
- **중재**

중재 및 비교 중재에 대한 상세한 설명
  - 구성 요소, 전달 경로, 투여량, 시점, 빈도, 중재 프로토콜, 중재 기간
  - 실행 관련 요소(예 : 연구원 자격, 장비 요구 사항)
  - 중재 진실성(즉, 중재의 특정 절차, 구성 요소가 계획대로 이행된 정도)
  - 공동 중재(Co-intervention) 기술
  - '대조군'의 정의(예 : 중재 없음, 위약, 최소 활동 비교군 또는 일반적인 치료 요소)
  - 연구자 또는 체계적 문헌고찰 저자의 기타 의견
  - 구성 요소, 용량, 시간, 빈도
  - 관찰 연구의 경우 : 중재 상태가 어떻게 평가되었는지; 노출 기간, 누적 노출 시간

• **중재결과(Outcomes)**

체계적 문헌고찰에서 각 중재결과 영역에 대해

- 중재결과 영역이 평가되었다는 근거가 있는지 여부(중재결과가 평가되었지만 연구결과가 제시되지 않은 경우 특히 중요)
- 측정도구: 척도 이름(예 : Hamilton Anxiety Rating Scale), 상한선 및 하한선, 높은 점수 또는 낮은 점수가 유리한지 여부, 임계치에 대한 정의
- 특정 메트릭(예 : 중재 후 불안, 기저 상태에서 중재 후 시간까지 불안 변화, 중재 후 불안 유무(예 / 아니오)
- 집계 방법(예 : 각 그룹의 불안 점수의 평균 및 표준 편차, 또는 불안이 있는 사람들의 비율)
- 중재결과 측정 시점(예 : 8주 중재 기간 종료시 평가, 8주 중재 기간 중 발생한 사건)
- 유해한중재결과는 체계적으로 또는 비체계적(예 : 자발적 보고)으로 수집되는지 여부

• **연구결과**

- 각 그룹 및 각 시점에서 중재결과에 대해 : 분석에 배정, 등록, 포함된 연구대상자 수; 중단한 연구대상자 수, 추적을 못했거나 제외된 연구대상자 수(각 이유와 함께)
- 각 그룹의 요약 자료(예 : 이분형 자료의 2x2 테이블, 연속형 자료의 평균 및 표준 편차)
- 중재가 중재결과에 미치는 영향을 정량화하는 집단 간 추정치와 정밀성(예 : 위험비, 오즈비, 평균 차이)
- 하위군 분석이 계획되어 있는 경우 각 연구대상자 군에 대해 동일한 정보를 추출해야 함

• **기타**

- 연구자의 주요 결론
- 기타 관련 연구의 참고문헌

5.3.2 연구방법과 비뚤림에 대한 출처

다른 연구방법에 따라 비뚤림 정도가 달라져 중재결과도 달라질 수 있다. 중요한 연구설계 특성을 수집하여 평가 및 분석을 위한 적절한 방법을 선택하고 포함된 각 연구의 설계를 '포함된 연구의 특성표'에 기술 할 수 있도록 해야 한다. 비 무작위배정 임상시험을 포함한다면, 연구의 적절한 특성을 기술해야 한다.

포함된 각 연구에서 비뚤림위험을 평가하기 위해 자세한 정보를 수집해야 한다. 비뚤림위험은 각 연구의 설계에 가장 적합한 도구를 사용하여 수행해야 하며 평가를 완료하는 데 필요한 정보는 도구에 따라 다르다.

연구 자금 출처와 연구 저자의 잠재적인 이해 상충도 중요한 정보이다.

일부 체계적 문헌고찰 저자는 연구 수행의 질에 영향을 미치지만 직접적인 위험을 초래하지

는 않는 연구 특성에 대한 추가 정보(예 : 윤리적 승인을 얻었는지 여부, 표본 크기 계산을 수행했는지 여부)를 수집하기도 한다.

5.3.3 연구대상자와 세팅

연구대상자의 세부 사항은 포함된 연구대상자 간의 비교 및 차이점을 이해하고, 포함된 연구대상자가 문헌고찰 질문을 적절하게 반영하는지를 평가할 수 있도록 수집된다.

연구대상자의 특성 중 중재 효과 여부나 크기에 영향을 미칠 수 있는 요소와 문헌고찰에서 다루지 않는 인구집단에 대한 적용 가능성을 평가하는 데 도움이 될 수 있는 요소를 수집해야 한다. 예를 들어, 체계적 문헌고찰 저자가 다른 사회경제적 그룹 간에 중재 효과에 중요한 차이가 있다고 의심되는 경우 이 정보를 수집해야 한다. 그러나 그러한 집단에 대해 중재 효과가 일정하다고 생각될 때 수집해서는 안 된다. 적용 가능성을 평가하는 데 유용한 연구대상자 특성에는 연령 및 성별이 포함된다. 이러한 특성은 여러 가지 다른 형식(예 : 평균 또는 중간 값, 표준 편차 또는 범위, 전체 또는 각 중재군별 성별 백분율 등)으로 표시될 수 있다. 체계적 문헌고찰 저자는 가능하면 전체 또는 중재군별로 특성을 요약하는 것 중 무엇이 더 적절한지 결정해야 한다. 때로 인종, 사회 인구학적 세부 사항(예 : 교육 수준) 및 병력 여부 같은 특성을 포함할 수 있다. 문헌고찰 질문과 관련된 임상 특성(예 : 당뇨병에 대한 문헌고찰의 경우 혈당 수준)은 또한 질병의 심각성 또는 단계를 이해하는 데 중요하다.

관심 대상 질병 상태를 정의하는 데 사용된 진단 기준은 특히 중요하다. 예를 들어, 울혈성 심부전의 약물치료에 대한 문헌고찰에서 심부전의 정의와 심각성이 수축기 또는 이완기 기능 장애, 20% 미만의 박출 계수(ejcetion fraction)를 가진 중증의 수축기 기능 장애 등 각 연구에서 어떤 진단기준을 사용하였는지를 아는 것이 중요하다. 마찬가지로, 항고혈압 치료에 대한 문헌고찰에서는 연구대상자의 혈압의 기저 수준을 설명하는 것이 중요하다.

연구의 세팅이 중재 효과 또는 적용 가능성에 영향을 줄 수 있다면, 이들에 대한 정보를 수집해야 한다. 보건의료 중재 연구의 일반적인 세팅에는 급성 치료 병원, 응급 치료 시설, 일반 진료시설 및 노인요양시설, 외래, 학교 및 지역사회와 같은 장기치료시설이 포함된다. 때로는 연구가 문화적 특성, 경제적 맥락, 시골과 도시의 환경 세팅과 같은 중재 및 중재결과에 영향을 미칠 수 있는 중요한 차이점을 지닌 다른 지역에서 실시되기도 한다. 연구의 시점은 중요한 의료기술 차이 또는 시간 경과에 따른 추세와 관련될 수 있다. 그러한 정보가 문헌고찰의 해석에 중요하다면, 그 정보가 수집되어야 한다.

포함된 각 연구의 연구대상자의 중요한 특성은 '포함된 연구의 특성' 표에서 요약되어야 한다.

5.3.4 중재

문헌고찰과 관련된 모든 실험 및 비교 중재의 세부 사항을 수집해야 한다. 즉, 효과 여부 또는 크기에 영향을 미칠 수 있거나 사용자의 상황에 따른 적용 가능성을 평가하는 문헌고찰에서는 세부 정보가 필요하다.

TIDieR(The Template for Intervention Description and Replication)은 중재에 대한 포괄적인 개념적 틀을 제공하며 체계적 문헌고찰 뿐 혹은 일차연구 보고에 사용하기 위해 제안되었다. 체크리스트에는 다음 내용이 포함된다.

- 중재 사용 근거와 어떻게 작동할 것으로 예상하는지
- 중재 대상자를 위한 중재 안내문서
- 중재를 제공하기 위해 수행하는 작업(절차 및 과정)
- 중재를 제공하는 사람(기술 수준을 포함), 방법(예 : 직접 대면, 웹 기반) 및 세팅(예 : 집, 학교 또는 병원)
- 시점과 강도
- 어떠한 변이(variation)가 허용 또는 예상되는지, 그리고 실제 변경(modification)이 이루어졌는지 여부
- 중재에 대한 충실도 또는 준수를 보장하거나 평가하기 위해 사용된 모든 전략, 그리고 중재가 계획대로 전달된 정도

의약품 임상시험의 경우, 수집할 핵심 정보에는 투여 경로(예 : 경구 또는 정맥 내 투여), 투여량(예 : 각 치료의 양 또는 강도, 투여 빈도), 시점(예 : 진단 후 24 시간 이내) 및 치료 기간 등이 있다. 심리 요법, 행동 및 교육 접근법, 또는 의료 전달 전략을 평가하는 것과 같은 경우 중재의 특성을 나타내는 데 필요한 정보의 양이 일반적으로 더 많을 것이다. 포함된 각 연구에서 중재의 중요한 특징은 '포함된 연구의 특성' 표에 요약되어야 한다. 로직 모델과 같은 추가 표나 그림은 문헌고찰 사용자가 자신의 상황에 대한 문헌고찰의 적용 가능성을 더 잘 평가할 수 있도록 해준다.

5.3.3.1 중재 진실성(Integrity)

특정 절차나 중재 요소가 계획대로 이행되는 정도는 연구결과에 중요한 영향을 미칠 수 있다. 우리는 이것을 중재 진실성이라고 부른다. 관련 용어에는 준수, 충실도 등이 있다. 중재 진실성의 검증은 행동 중재와 복잡한 중재 같은 비약물 임상시험 문헌고찰에서 특히 중요 할 수 있는데, 이는 이상적인 중재가 제공되기에는 많은 장애가 있는 상태에서 실행되기 때문이다.

중재가 계획대로 실행되었는지 여부를 평가할 때 문헌고찰 저자는 일부 중재는 표준화되고(중재 프로토콜에서 허용되는 편차 없음), 일부 중재는 어느 정도 맞춤을 허용한다는 점을 명심해야 한다. 진실성에 대한 정보는 불만족스러운 중재결과가 잘못 개념화된 중재로 인한 것인지 아니면 규정된 구성 요소의 불완전한 제공으로 인한 것인지를 판단하는 데 도움이 될 수

있다. 또한 실제 환경에서 주어진 중재의 실제 타당성에 대한 중요한 정보를 줄 수 있다.

5.3.3.2 과정평가

과정 평가는 중재의 의도된 실행과 실제 효과 사이의 과정을 평가하는 것이다. 과정 평가 연구들은 자료수집에 대한 유연한 접근과 양적 및 질적 방법을 모두 포함하는 다양한 유형의 자료를 생성하는 다양한 방법의 사용을 특징으로 한다. 핵심 과정 요소를 설명했는지 또는 측정했는지 여부와 진실성을 철저하게 다루었는지가 중요하다. 과정 평가는 중재의 효과에 잠재적으로 영향을 미치는 요소에 유용한 출처가 될 수 있다.

5.3.5 중재결과

중재결과는 중재 중 또는 중재 후 연구에서 특정 개인 또는 중재 단위에서 관찰되거나 기록된 사건 또는 측정값이며, 조사된 중재의 효과 및 안전성을 평가하는 데 사용된다. 체계적 문헌고찰 저자는 연구에서 측정된 모든 중재결과에 대한 정보를 수집할지 아니면 관심 이 있는 중재결과만 수집할지를 미리 지정해야 한다.

프로토콜에 정의된 중재결과만 자세히 설명하는 것이 좋다. 그러나 측정된 모든 중재결과의 전체 목록은 결측된 중재결과 자료로 인한 비뚤림의 위험을 보다 자세하게 평가할 수 있게 한다.

문헌고찰 저자는 중재결과의 다섯 요소를 모두 수집해야 한다.

1. 중재결과 영역 또는 이름(예 : 불안)
2. 측정 도구(임상적 중재결과 또는 종결점의 정의 포함); 척도 이름, 척도의 상한선 및 하한선, 높거나 낮은 점수가 좋은 것인지 여부, 임계치의 정의
3. 연구대상자의 연구결과를 특성화하는데 사용된 메트릭(예 : 중재 후 불안 정도, 중재에 의한 불안 정도의 변화, 중재 후 불안 유무 등)
4. 집계 방법(예 : 각 그룹의 불안 점수의 평균 및 표준 편차, 불안이 있는 사람들의 비율)
5. 중재결과 측정의 시점(예 : 8주 중재 기간 종료 시 평가, 8주 중재 기간 중 발생한 사건)

5.3.5.1 유해효과

중재의 위해한 영향의중재결과는 유해사건(adverse event), 유해효과(adverse effec), 약물유해반응(adverse drug reaction), 부효과(side effect), 합병증(complication) 등으로 기술된다. 이러한 전문 용어는 기술적으로는 다를 수 있지만, 문헌상에서는 서로 바꿔가며 사용된다. 위해(harms)는 삶의 질의 개선이 예상되는 상황에서 삶의 질이 저하되는 것과 같이, 연구 중에 측정된 중재결과에서 바람직하지 않은 변화를 포함하는 것으로 해석될 수 있다.

임상시험에서 유해효과는 체계적으로 또는 비체계적으로 수집할 수 있다. 체계적 수집은 설문지 또는 실험실 검사와 같이 정의된 방법을 사용하여 각 연구대상자에게 동일한 방식으로 유해사건을 수집하는 것을 의미한다. 비체계적 수집은 개방형 질문(예 : '지난 방문 이후 증상

이 나타났습니까?')과 같은 비체계적 방법을 사용하여 유해사건에 대한 정보를 수집하거나 연구대상자가 자발적으로 보고하는 것을 말한다. 두 경우 모두 유해효과는 중증도에 따라 선택적으로 보고될 수 있으며, 연구대상자는 그 영향이 중재에 의해 유발된 것인지 여부를 판단하기 때문에 비뚤림이 생길 수 있다. 다음은 유해효과에 대해 수집할 유용한 정보이다.

- 버전 번호를 포함하여 사용된 모든 코딩 시스템 또는 표준 의학 용어(예 : COSTART, MedDRA)
- 유해사건 이름(예 : 현기증)
- 보고된 유해사건의 강도(예 : 경증, 중등도, 중증)
- 임상시험 연구자가 유해사건을 '심각'으로 분류했는지 여부
- 임상시험 연구자가 이 유해사건을 중재와 관련이 있다고 확인했는지 여부
- 시간 프레임(일반적으로 연구 기간 동안 계수로 측정)
- 어느 정도의 유해사건을 보고 했는지(예 : 적어도 5%의 연구대상자에서 발생한 모든 유해사건을 보고함)
- 관련 연구결과

비체계적 수집 방법은 유해사건이 얼마나 자주 발생하는지 과소평가하는 경향이 있다. 또한, 유해사건이 체계적이지 않은 정량적 선택 기준에 근거하여 보고된 경우(예 : 연구대상자의 최소 5%에서 발생한 유해사건만이 출판물에 포함됨), 보고된 자료만으로는 메타분석의중재결과에 비뚤림이 될 수 있다. 체계적 문헌고찰 저자는 다르게 수집된 유해사건을 합성하는 것에 주의해야 한다. 수집 방법에 관계없이, 유해효과와 그 강도의 정확한 정의는 연구마다 다를 수 있으므로 기록해야 한다.

임상시험 보고서에서 유해사건에 대한 언급이 없다고 해서 유해사건이 발생하지 않았다는 것을 반드시 의미하는 것은 아니다. 일반적으로 보고되지 않았다고 가정하는 것이 가장 안전하다.

5.3.6 연구결과

연구결과 자료는 중재연구에서 개별 연구대상자의 중재결과 측정 또는 확인으로 생성된다. 연구결과는 각 개인 또는 군 수준(arm level)에서 요약되거나 두 중재군을 비교함으로써 중재효과가 요약될 수 있다. 연구결과 자료는 프로토콜에 지정된 중재군과 중재결과에 대해서만 수집해야 한다. 다른 중재결과에 대한 연구결과는 프로토콜을 수정하지 않는 한 수집되어서는 안 된다. 그러나 문헌고찰 저자는 중요한 연구결과, 예상치 못한 결과, 특히 중대한 유해효과의 가능성에 주의해야 한다.

연구 보고서에는 종종 동일한 중재결과지표에 대한 몇 가지 연구결과가 포함된다. 예를 들어, 서로 다른 측정 척도가 사용될 수 있으며, 결과는 여러 하위군에 대해 개별적으로 제시될

수 있고, 연구결과는 다른 시점에서 측정되었을 수 있다. 틀(Framework)은 여러 측정 또는 연구결과 사이에서 선택을 용이하게 하는 효과가 있으며, 프로토콜에 미리 구체화되어야 한다. 일단 자료를 수집한 이후 이 틀에 대해 추가적인 결정을 하거나 변경하는 경우 해당 내용을 프로토콜 변경으로 문헌고찰시 보고되어야 한다.

중재결과 자료의 유형은 각 중재결과에 대해 조사할 수치의 성격을 결정한다. 예를 들어, 이분형('예' 또는 '아니요') 중재결과의 경우 연구대상자 수와 중재결과를 경험한 사람 수를 각 군별로 제시해야 한다. CONSORT에서 권장하는 흐름도(flow diagram)는 연구과정에서 연구대상자의 흐름을 결정하는 데 도움이 될 수 있다. 출판된 보고서에 흐름도가 없는 경우 체계적 문헌고찰 저자는 그림을 그리는 것을 고려할 수 있다(www.consortstatement.org에서 제공).

메타분석에 필요한 수치자료가 항상 사용 가능한 것은 아니다. 흔히 다른 통계를 수집하여 필요한 형식으로 변환할 수 있다. 예를 들어, 연속형 중재결과를 얻으려면 일반적으로 연구대상자 수, 각 중재군의 평균 및 표준 편차를 찾는 것이 가장 편리하다. 이들은 종종 직접적으로 이용 가능하지 않으며, 표준편차가 이용가능하지 않은 경우가 많다. 대안적인 방법으로 표준오차, 신뢰 구간, 검정통계(예 : t- 검정 또는 F- 검정) 또는 p값으로 결측된 표준 편차를 계산하거나 추정할 수 있다. 이러한 자료가 유용한 정보를 제공한다면 추출해야 한다.

5.3.7 수집할 기타 정보

체계적 문헌고찰 저자는 저자가 보고한 대로 포함된 연구의 주요 결론을 수집하는 것이 좋다. 문헌고찰시 이러한 결론을 보고 할 필요는 없지만, 체계적 문헌고찰 결과에서 효과의 방향과 관련하여 특히 문헌고찰 저자의 분석 결과를 검증하는 데 사용되어야 한다. 예기치 못한 연구결과에 대한 설명과 같은 연구 저자의 추가 의견을 언급할 수도 있다.

5.4 자료 수집 도구

5.4.1 자료수집 양식에 대한 근거

체계적 문헌고찰을 위한 자료수집은 구조화된 자료수집 양식을 사용하여 수행해야 한다. 자료수집 양식에는 종이 양식, 전자 양식(예 : Google 양식) 또는 상업용 또는 맞춤형 자료 시스템(예 : Covidence, EPPI-REVIEWER, Systematic Review Data Repository(SRDR)이 있다. 맞춤형 자료 시스템은 온라인 양식 작성을 허용하며, 자료 공유 및 효율적인 자료관리를 가능하게 한다.

자료 수집 양식은 연구자가 보고한 내용과 체계적 문헌고찰 저자가 궁극적으로 보고하는 내용 사이의 다리 역할을 한다. 자료수집 양식은 몇 가지 중요한 기능을 제공한다. 첫째, 자료수

집 양식은 문헌고찰 질문 및 적격기준에 직접 연결되어 있으며, 그런 것들에 대한 명확한 요약을 제시해서 연구 보고서에서 추출할 자료를 식별하고 구조화하는 데 도움을 준다. 둘째, 문헌고찰 과정에서 발생하는 수많은 의사 결정뿐만 아니라 문헌고찰 과정에서 사용된 자료의 출처에 대한 기록이다. 셋째, 양식은 분석에 포함할 자료의 출처이다.

자료수집 양식의 중요한 기능을 고려할 때 충분한 시간과 생각을 설계에 투자해야 한다. 모든 문헌고찰이 다르므로 자료수집 양식도 문헌고찰마다 다르다. 그러나 중요한 정보 유형에는 많은 유사점이 있다. 따라서 양식을 다른 문헌고찰에서 채택해 문헌고찰에 적용할 수 있다. '자료수집 양식'이라는 용어를 사용하기는 하지만 실제로는 다른 목적으로 사용되는 일련의 형식이 될 수 있다. 문헌고찰에서 제외되거나 문헌고찰에 포함될 연구의 신속한 식별에 사용된다.

5.4.2 자료수집 도구 선택 시 고려사항

자료수집 도구의 선택은 체계적 문헌고찰 저자의 선호도, 문헌고찰 규모 및 저자가 사용할 수 있는 자원에 주로 좌우된다. 종이양식(워드 프로세싱 소프트웨어를 사용하여 개발된 양식)은 동일한 장소에 2-3명의 자료 추출자가 있는 소규모 팀에 적절하고 메모와 설명을 쉽게 기록 할 수 있다. 그러나 소프트웨어에 자료를 입력해야 하므로 비효율적이며 에러에 취약하다. 전자양식(Microsoft Access, Google Forms)은 중소 규모의(4-6명의 자료 추출자) 팀에 적절하며 추출된 자료가 편집과 분석을 위해 전자적으로 처리되고, 자료 저장, 공유 및 자료 정렬이 가능하고, 필요에 따라 양식을 쉽게 확장하거나 편집할 수 있다. 그러나 소프트웨어 버전의 변경에 취약하다. 자료 시스템(Covidence, EPPI-Reviewer, Systematic Review Data Repository, Distiller Systematic Review(Evidence Partners) Doctor Evidence)은 특히 대규모 팀(6명 이상의 자료 추출자)에 적합하며, 온라인에서 자료를 저장, 연결, 공유가 가능하고 필요에 따라 양식을 쉽게 확장하거나 편집할 수 있다. 또한 동일 연구에 대한 독립적인 자료수집 간의 자료 비교가 가능하고, 팀 진행 상황을 모니터링 할 수 있으며, 여러 저자가 동시에 입력이 가능하다. 또한 자료를 분석 소프트웨어로 직접 내보낼 수 있다. 그러나 양식을 세팅하고 자료 추출자를 교육하기 위한 선행 투자가 필요하며, 상업 자료 시스템의 경우 비용이 들고, 소프트웨어 버전 변경에 취약하다.

5.4.3 자료수집 양식 설계

성공적인 자료수집의 핵심은 사용하기 쉬운 양식을 만들고 구조화되고 체계적인 방법으로 충분하고 모호하지 않은 자료를 수집하는 것이다. 대부분은 전자 양식이나 자료 시스템을 만들기 전에 양식에 대한 문서 형식을 개발해야 한다. 이것은 프로그래머 및 자료 분석가를 비롯한 다른 사람들에게 배포될 수 있으며 전자 양식 및 자료 추출자가 사용할 지침이나 코드북을

작성하는 지침으로도 사용할 수 있다. 체계적 문헌고찰 저자는 전자 양식이나 자료 시스템과 분석 소프트웨어의 호환성 및 자료 입력 오류를 기록, 평가 및 수정하는 메커니즘도 고려해야 한다.

체계적 문헌고찰 저자는 각 개별 연구에 대해 여러 보고서가 이용 가능한 경우 어떤 보고서를 우선시하여 자료를 추출할 것인가를 프로토콜에 명시하여야 한다. 예를 들어 연구자에게 연락하는 방법으로 해결할 수 없는 자료 충돌이 있을 때 우선시할 보고서를 미리 정해야 한다. 마찬가지로 연구에 대해 하나의 보고서만 있는 경우도 섹션별 불일치 자료가 있는 경우 어디에서 자료를 추출할지(예 : 요약, 방법, 결과, 표 및 그림)를 지정해야 한다.

좋은 자료수집 양식은 원본 문서로 돌아갈 필요를 최소화해야 한다. 자료수집 양식을 설계할 때 팀의 모든 구성원, 즉 내용 영역 전문가, 체계적 문헌고찰 방법 및 자료수집 양식 설계에 경험이 있는 저자, 통계 전문가의 도움을 받아야 한다. 자료수집 양식을 설계하기 위한 다음의 몇 가지 팁이 있다.

- 1단계. 문헌고찰 과정에서 측정할 다양한 중재결과를 고려하여 체계적 문헌고찰에 나타날 것으로 예상되는 표와 수치 자료의 개요를 개발한다. 이 단계는 저자가 수집할 자료의 양(너무 많지 않거나 너무 적지 않음)을 정하는데 도움이 된다. 너무 많은 정보를 수집하면 원래의 연구 보고서보다 긴 양식으로 이어질 수 있으며 시간 낭비가 될 수 있다. 너무 적은 정보를 수집하거나 주요 자료가 결측되면 문헌고찰 과정의 후반부에 연구 보고서로 돌아갈 필요가 생길 수 있다.
- 2단계. 양식 개발을 용이하게 하기 위해 자료 요소를 모으고 그룹화한다. 자료 요소를 연구 보고서에서 일반적으로 발견되는 순서(예 : 참고문헌 정보, 적합성 기준, 중재 설명, 통계 방법, 기저상태 특성 및 결과)로 그룹화 하는 것이 더 효율적일 수 있다.
- 3단계. 자료 항목 형태화를 위한 최적의 방법을 확인한다. 다음과 같은 것이 있다.

  1. 가능한 한 폐쇄형 질문을 한다.
  2. 응답을 비워 둘 수 있는 방식으로 질문하는 것을 피한다. 필요에 따라 '해당 사항 없음', '보고되지 않음' 및 '말할 수 없음'을 포함한다.
  3. 이 양식은 연구에서 수행된 것보다 논문에 보고된 것에 초점을 맞춰야 한다.
  4. 판단이 필요한 경우, 판단에 사용된 원시 자료(즉, 원본 문서의 출처)를 기록한다.
  5. 자료 보고의 변이를 수용하는 유연성을 갖는다. 필요한 경우 후속 단계에서 결과 자료를 보고하고, 변환한 형식으로 결과 자료를 수집하는 것이 좋다.

- 4 단계. 자료 수집 양식을 개발하고 파일럿 테스트하여 이후 분석을 위해 올바른 형식 및 구조로 자료를 제공하도록 한다. 자료수집 양식에는 2단계에서 설명한 자료 항목 외에 작성자, 완료일, 문헌고찰 제목도 기록해야 한다. 양식에 간혹 개정이 필요하다.

모든 자료수집 양식과 자료 시스템은 출시 전에 철저한 파일럿 테스트를 거쳐야 한다. 테스트에는 적어도 몇 가지 논문에 대해 여러 자료 추출자가 동시에 추출을 시행한다. 초기 테스트

는 질문의 명확성과 완성도에 중점을 둔다. 테스트는 양식에서 결측이 되었거나 불필요한 자료를 식별할 수 있다. 초기 테스트 후 추출된 자료의 정확성을 원본 문서 또는 확인된 자료와 비교하여 문제가 있는 영역을 확인해야 한다. '포함된 연구의 특성표'에 대한 항목 초안을 마련하고 모든 필요한 정보가 수집되도록 이러한 파일럿 보고서를 사용하여 비뚤림 평가의 위험을 완수하는 것이 좋다. 나중에 불일치를 피하기 위해 양식을 수정하기 전에 저자 간의 합의가 필요할 수 있다. 첫 번째 파일럿 테스트 후에 큰 변경이 필요한 경우 파일럿 테스트를 반복해야 할 수도 있다.

파일럿 테스트가 완료된 후에도 자료수집 양식의 문제가 발생할 수 있으며 자료 추출이 시작된 후 양식을 수정해야 할 수도 있다. 양식 또는 코딩 지침이 변경되면 이미 자료 추출이 완료된 보고서로 돌아가야 할 수도 있다. 경우에 따라 실제 자료수집 양식을 수정하지 않고 코딩 지침만 명확히 해야 할 수도 있다.

5.5 보고서에서 자료 추출

5.5.1 소개

대부분의 체계적 문헌고찰에서 각 연구에 대한 주요 정보 출처는 일반적으로 학술지에 발표된 연구 보고서이다. 텍스트를 전자 검색하면 보고서 내에서 정보를 찾는 데 도움이 된다. 예를 들면 PDF 뷰어, 인터넷 브라우저 및 워드 프로세싱 소프트웨어에서 검색 기능을 사용하는 경우가 있다. 그러나 정보는 다양한 용어로 제시될 수 있고 여러 형식으로 제시되기 때문에, 보고서를 읽는 대신에 텍스트 검색으로 대체할 수는 없다.

5.5.2 자료 추출자 결정

자료 추출자는 주제에 대한 기본적인 이해가 있어야 하며 연구설계, 자료 분석 및 통계에 대한 지식을 갖추고 있어야 한다. 주관적 해석과 결과의 해석이 중요한 정보가 포함된 정보는 최소 두 사람이 독립적으로 추출해야 한다. 자료 추출자는 방법론 전문가 혹은 주제 분야 전문가와 같이 보완적인 분야에서 뽑는 것이 바람직하다.

5.5.3 자료 추출자 훈련

자료 추출자 교육은 문헌고찰 주제 및 방법, 자료수집 양식 또는 자료 시스템 및 자료 추출 중에 발생할 수 있는 문제에 익숙해지도록 하는 것이 목적이다. 양식의 파일럿 테스트 결과를 보면서 모호한 질문이나 응답이 있으면 문헌고찰 저자와 추출자들이 토론을 통해 해결한다. 교육은 자료 추출 과정이 시작될 때 그리고 프로젝트 진행 과정 중에 주기적으로 수행되어야 한다.

5.5.4 동일한 연구에 대한 여러 보고서에서 자료 추출

연구는 하나 이상의 출판물 또는 하나 이상의 출처에서 보고된다. 여러 출처는 동일한 연구에 대한 상반되는 정보를 포함할 수 있다. 체계적 문헌고찰에서 관심 단위는 보고서가 아니라 연구이기 때문에 여러 보고서의 정보를 대조하고 조정해야 하는 경우가 있다. 기본 보고서에 포함되지 않은 귀중한 정보가 포함되어 있을 수 있으므로 포함된 연구 보고서를 면밀히 검토하지 않고 버리는 것은 적절하지 않다. 체계적 문헌고찰 저자는 두 가지 전략 중 하나를 결정해야 한다.

- 각 보고서에서 별도로 자료를 추출한 다음 여러 자료수집 양식에서 정보를 결합.
- 모든 보고서의 자료를 단일 자료수집 양식으로 직접 추출.

사용할 전략의 선택은 보고서의 성격에 따라 다르며 연구 및 보고서에 따라 다를 수 있다. 예를 들어, 전체 학술지 논문과 여러 학술대회 초록을 사용할 수 있는 경우 대부분의 정보는 학술지 논문에서 얻을 수 있다. 각 학술대회 요약을 위한 새로운 자료수집 양식을 작성하는 것은 시간 낭비일 수 있다. 반대로 후속 작업 기간이 서로 다른 두 개 이상의 상세 학술지 논문이 있는 경우 해당 논문에 대해 별도로 자료 추출을 수행하고 나중에 자료수집 양식의 정보를 대조하는 것이 더 쉬울 수 있다. PRISMA와 같은 흐름도는 여러 보고서의 정보를 대조하고 문서화 할 때 특히 유용할 수 있다.

5.5.5 신뢰도와 합의 도달

자료가 두 명 이상의 추출자에 의해 독립적으로 추출된 후에는 불일치를 확인하기 위해 응답을 비교해야 한다. 불일치를 식별하고 해결하기 위해 프로토콜에 절차 또는 의사결정 규칙을 명시해야 한다. 불일치 해결은 저자 간의 토론이 합리적인 첫 단계이다. 의견 차이로 다른 사람이 중재해야 할 수도 있다. 해결할 수 없는 의견 불일치는 연구 저자에게 연락하여 해결해야 하고, 해결되지 못한 경우 문헌고찰에서 불일치 사항을 보고해야 한다. 불일치의 존재와 해결은 신중하게 기록되어야 한다. 합의된 자료 이외에 '추출된' 자료의 복사본을 유지하면 코딩의 신뢰도를 평가할 수 있다. 불일치 합의 방법의 예는 다음과 같다.

- 한 사람의(종이) 자료수집 양식을 사용하여 합의 후에 다른 색으로 변경 사항을 기록함.
- 전자 양식에 합의 자료를 입력한다.
- 추출된 원본 자료와 합의 자료를 별도의 형식으로 기록한다(일부 온라인 도구는 자동으로 이 작업을 수행함).

5.5.6 임상연구 보고서에서 자료 추출

임상연구 보고서는 길이가 수천 페이지에 달하고 문헌고찰 하는데 시간이 오래 걸리지만, 정보는 일반적으로 체계적이고 논리적인 방식으로 제공된다. 임상연구 보고서에서 제공되는 정보의 명확성과 완전성으로 인해 자료 추출은 학술지 논문이나 기타 짧은 보고서보다 분명하고 명확하게 수행될 수 있다.

임상연구 보고서에서 자료를 효율적으로 추출하려면 저자는 임상연구 보고서 구조에 익숙해야 한다. 실제로 문헌고찰 저자는 섹션 제목 및 하위 제목을 기록하고 자료 추출과 관련된 키워드(예 : 무작위 추출)를 검색하거나 북마크'를 만들 수 있다. 또한 추가 분석이 필요할 때 광학 문자 인식 소프트웨어를 사용하여 PDF 자료 표를 분석 가능한 형식으로 변환하여 시간을 절약하고 인쇄 오류를 최소화 할 수 있다.

임상연구 보고서는 평가되었으나 일반에게 공개되지 않은 결과를 확인하는 데 특히 유용하다. 효능 결과 및 체계적으로 유해효과를 수집하기 위해 저자는 보고서의 내용을 비교하여 결과 자료의 결측으로 인한 비뚤림위험을 평가하고 보고된 결과를 선택한다.

5.5.7 규제 심사자료에서 자료 추출

체계적 문헌고찰과 가장 관련이 있는 자료는 규제문서에 있는 의학 및 통계 심사자료 섹션에서 확인할 수 있다. 이 두 가지 모두 학술지 논문보다 훨씬 길다. 의약품에 대한 모든 임상 임상시험 목록은 보통 의학 심사자료에서 찾을 수 있다. 임상시험은 숫자와 문자의 조합으로 명명되기 때문에 여러 임상시험 보고서를 연결하는 것이 어려울 수 있다.

광학 문자 인식 소프트웨어는 대부분의 텍스트를 변환할 수 있다. 최신 의약품에 대한 심사자료가 문서에서 민감한 정보를 삭제하고 제공되어 검색이 가능하다. 임상연구 보고서와 비교하여 규제 심사자료에는 임상시험 설계, 실행 및 결과에 대한 정보가 적어 비뚤림위험을 평가하기 위한 정보는 제한적이다.

5.5.8 소프트웨어를 이용한 그림에서 자료 추출

때로는 체계적 문헌고찰에 필요한 수치 자료가 그림으로만 표시된다. 체계적 문헌고찰 저자는 연구자에게 자료를 요청하거나 수동 또는 소프트웨어를 사용하여 그림에서 자료를 추출할 수 있다. 많은 도구가 무료이다. Plot Digitizer, WebPlotDigitizer, Engauge, Dexter, ycasd, GetData Graph Digitizer라는 도구가 있다. 소프트웨어는 그림의 이미지를 찍은 다음 사용자가 세팅한 축과 눈금을 사용하여 그림에 있는 정보를 수치화한다. 정보를 체계적 문헌고찰에 사용할 수는 있지만, 개개인 수준의 자료 포인트로부터의 평균 및 표준 편차와 같은 필요한 요약 통계를 얻으려면 추가 계산이 필요할 수 있다.

5.5.9 체계적 문헌고찰에서 자료 추출 자동화

자료 추출은 시간이 오래 걸리고 오류가 발생하기 쉽기 때문에 이 단계를 자동화 또는 반자동으로 실행하면 추출 과정을 보다 효율적이고 정확하게 수행할 수 있다. 하지만 아직은 정기적인 체계적 문헌고찰 생산을 위해 자료 추출을 자동화하는 특정 도구를 추천할 수 없다.

5.5.10 과학적 부정행위에 대한 의혹

연구부정행위는 연구결과의 변조 위조, 표절 등 윤리적 규범을 준수하지 않는 것을 말한다. 체계적 문헌고찰 저자는 문헌고찰 결과의 신뢰도를 저해할 수 있는 과학적 부정행위에 대해 알고 있어야 한다. 중복 출판물의 경우 합성 결과가 두 배로 계산될 수 있다.

조사에는 시간이 걸릴 수 있기 때문에 부정행위가 의심되는 연구는 '평가 대기 중'으로 분류해야 한다. 부정행위 조사 결과 출판물을 신뢰할 수 없거나 취소된 경우 체계적 문헌고찰에 포함되어서는 안 되며 배제 이유는 '제외된 연구' 섹션에 기록되어야 한다.

5.5.11 자료 추출 계획 및 보고의 요점

프로토콜과 체계적 문헌고찰의 방법에서는 다음을 상세하게 설명해야 한다.

- 추출할 자료 범주
- 각 보고서에서 추출된 자료를 검증하는 방법(예 : 두 명이 독립적으로 추출)
- 자료 추출이 내용 전문가, 방법론 전문가 또는 둘 다에 의해 수행되는지 여부
- 자료수집 양식에 대한 코딩 지침, 파일럿 테스트, 교육
- 동일한 연구의 여러 보고서에서 자료를 추출하는 방법
- 둘 이상의 저자가 각 보고서에서 자료를 추출할 때 불일치가 어떻게 처리하였는지

5.6 연구결과 추출 및 형식 변환

대부분의 경우 관심 중재군 별로 요약 자료를 수집하고 이를 RevMan과 같은 효과 추정치를 계산할 수 있는 소프트웨어에 입력하는 것이 바람직하다. 각 중재군에서 요약 자료를 얻을 수 없거나 보정된 분석 결과를 사용하는 것이 중요한 경우(예 : 교차 또는 군집무작위배정 임상시험에서 상관관계를 측정하기 위해) 효과 추정치를 직접적으로 사용할 수 있다.

5.7 자료관리 및 공유

개별 연구별로 자료를 수집한 경우, 자료를 메타분석 또는 기타 합성에 돌입하기 전에 데이

터베이스 또는 스프레드시트와 같은 포괄적인 전자 형식으로 바꾸는 것이 좋다. 자료를 전자 형식으로 바꾸면, 일관성 검사 및 분석을 하기 쉬워진다.

수집된 정보를 도표화하면 연구를 적절히 비교하고, 하위군으로 분류하는 것이 용이해질 수 있다. 또한 연구를 통해 비교 가능한 중재결과 측정치 및 통계치를 확인할 수 있다. 결과 제시나 합성을 위해 계산을 수행해야 하는 경우가 종종 있다. 이 과정을 통해 원본 문서의 자료와 계산을 통해 얻은 자료를 명확하게 구분하여, 자료의 출처에 대한 명확한 정보를 유지하는 것이 중요하다. 예를 들어, 표준 오차에서 표준 편차로 통계적 변환을 할 때는 실제 계산뿐만 아니라 원래 계산된 수의 영구적인 기록을 유지하기 위해 수동 계산기를 사용하는 것이 아니라 컴퓨터로 수행하는 것이 이상적이다.

이상적으로는 자료는 한 번 추출해야 하며, 문헌고찰의 향후 업데이트를 위해 안전한 위치에 저장해야 한다. 유사한 주제 영역에서 작업하는 체계적 문헌고찰 저자는 자료수집 시스템뿐 아니라 자료수집 도구를 표준화하고 공유하면 체계적 문헌고찰 생산을 간소화 할 수 있다. 체계적 문헌고찰 저자는 연구 자료(예 : 임상연구 보고서, 개별 연구대상자 자료 및 기타 다른 형태의 연구 자료)를 공개적으로 사용할 수 있도록 연구자, 학술지 논문 편집자, 후원자, 규제자 및 기타 이해 당사자와 협력하여 연구의 투명성을 높일 수 있다. 합법적이고 윤리적인 경우, 체계적 문헌고찰 저자는 낭비를 줄이고 확인 및 재분석을 허용하기 위해 체계적 문헌고찰에 사용된 자료를 공유하는 것이 좋다.

5.8 자료 수집 관련 MECIR

C48: 정정기사 검토(필수)	
관련 논문 취소에 대한 기술과 정정기사를 조사한다.	일부 연구는 사기성이 있거나 다른 이유로 출판 이후 논문취소가 되었을 수 있다. 정정기사는 포함된 연구에서 중요한 제한점이나 치명적인 결함을 포함할 수 있다. 이러한 연구는 모두 문헌고찰이나 메타분석에서 배제 시킬 수 있다. 이 정보는 인용 자료가 적절한 필드로 다운로드하여 모든 데이터베이스 검색에서 검색되도록 주의해야 한다.
C42: 여러 보고서 수집(필수)	
동일한 연구의 여러 보고서를 대조하여, 각 보고서보다는 연구가 문헌고찰의 관심 단위가 되게 한다.	동일한 연구에 대한 여러 보고서를 마치 여러 연구처럼 고려하는 것은 잘못된 것이다. 그러나 연구의 2차 보고서는 설계와 수행에 대한 중요한 정보를 포함할 수 있기 때문에 폐기해서는 안 된다. 체계적 문헌고찰 저자는 연구결과의 출처로 사용할 보고서를 선택하고 근거를 제시해야 한다.

C49: 미출판 자료 얻기(권고)	
포함된 연구의 보고서에서 미출판된 주요 미공개 정보를 찾는다.	자료를 얻거나 확인하기 위해 연구자에게 연락하면 문헌고찰이 보다 완벽해지고 잠재적으로 정밀도가 향상되고 보고 비뚤림의 영향이 줄어든다. 미출판 정보에는 비뚤림위험 평가, 중재 및 결과의 세부 사항 및 연구결과(중요한 하위군 별 결과의 분석 포함)에 대한 세부 정보가 포함된다.
C44: 연구 기술하기(필수)	
'포함된 연구의 특성표'를 채우기에 충분하고 상세하게 포함된 연구의 특성을 수집한다.	각 연구의 기본 특성은 연구대상자의 세부 사항, 중재와 비교중재, 중재결과 및 연구설계를 포함하여 문헌고찰의 일부로 제시되어야 한다.
C50: 여러 군(multi-arm) 연구에서 중재군 선택(필수)	
연구가 3개 이상의 중재군을 포함하는 경우, 문헌고찰에 적격 기준을 충족하는 중재군과 대조군만 포함한다.	문헌고찰에 관련 없는 중재군을 포함할 필요가 없다. 저자는 '포함된 연구의 특성표'에서 이러한 중재군이 연구에 있음을 분명히 해야 한다.
C47: 최대한 자료 활용(필수)	
포함된 연구와 유사한 분석을 할 수 있도록 가장 상세한 수치 자료를 수집하고 활용한다. 2×2 표 또는 평균과 표준 편차를 구할 수 없는 경우에는 효과 추정치(예 : 오즈비, 회귀 계수), 신뢰 구간, 임상시험 통계(예 : t, F, Z, 카이 제곱) 또는 p값 또는 개별 연구대상자 자료도 포함할 수 있다.	RevMan 자료 입력은 이분형 중재결과에 대해 2x2 표가 보고되고 연속형 중재결과에 대해 평균과 표준 편차가 제시될 때 가장 쉽다. 때때로 이러한 통계는 보고되지 않지만 계산할 수 있다. 예를 들어, 2x2 표는 표본 크기와 백분율에서 구할 수 있고, 표준 편차는 신뢰 구간이나 p값을 사용하여 계산할 수 있다. 또한, 역 분산 자료 입력 형식은 이분형 자료 또는 연속 자료에 필요한 세부 자료가 없는 경우에도 사용할 수 있다(예 : 오즈비 및 신뢰구간만 제시된 경우).
C43: 자료 수집 양식 사용(필수)	
예비평가 된 자료 수집 양식을 사용한다.	체계적 문헌고찰 저자들은 종종 경력이 다르며 체계적 문헌고찰 경험도 다르다. 자료 수집 양식을 사용하면 자료 추출 과정의 일관성을 보장할 수 있으며 중복된 자료를 비교하는 데 필요하다. 완성된 자료수집 양식은 요청시 CRG에 제출할 수 있어야 한다. 문헌고찰 팀 내에서 양식을 미리 테스트해보는 것이 매우 바람직하다. 자료수집 양식은 유용성을 평가해야 한다.
C45: 연구 특성을 중복하여 추출하는 것(권고)	
독립적으로 두 사람이 각 연구 보고서에서 연구 특성을 추출하고 불일치 해결 과정을 미리 정의한다.	자료 추출 과정을 중복하면 실수를 저지를 위험이 줄고 자료 선택이 한 사람의 편견에 영향을 받을 가능성이 줄어든다. 연구 특성에 대한 이중 자료 추출은 중재결과 자료보다는 덜 중요 하므로 필수 표준은 아니다.

C46: 중복 중재결과 자료 추출(필수)	
독립적으로 두 사람이 각 연구 보고서에서 중재결과 자료를 추출하고 불일치를 해결하는 과정을 미리 정의한다.	자료 추출 과정을 중복하면 실수를 저지를 위험이 줄고 자료 선택이 한 사람의 편견에 영향을 받을 가능성이 줄어든다. 중재결과 자료에 대한 이중 자료 추출이 특히 중요하며, 직접적으로 근거 합성과 문헌고찰 결론을 만들어 준다.

5.9 참고문헌과 읽을거리

- Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, Altman DG, Barbour V, Macdonald H, Johnston M, Lamb SE, Dixon-Woods M, McCulloch P, Wyatt JC, Chan AW, Michie S. Better reporting of interventions: template for intervention description and replication(TIDieR) checklist and guide. BMJ 2014;348: g1687.
- Li T, Higgins JPT, Deeks JJ(editors). Chapter 5: Collecting data. Draft version(29 January 2019) for inclusion in: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA(editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.0(updated July 2019). Cochrane, 2019. Available from www.training.cochrane.org/handbook.
- Li TJ, Vedula SS, Hadar N, Parkin C, Lau J, Dickersin K. Innovations in data collection, management, and archiving for systematic reviews. Annals of Internal Medicine 2015;162: 287-294.
- Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, 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 Medicine 2009;6: e1000100.
- Moher D, Schulz KF, Altman DG. The CONSORT Statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. Lancet 2001;357: 1191-1194.

제6장 비뚤림위험 평가의 개념과 무작위배정 임상연구 평가

핵심 포인트

- 비뚤림위험은 두 곳에서 고려해야 하는데 첫 번째는 체계적 문헌고찰에 포함된 개별 연구의 결과에서이고, 두 번째는 포함된 연구결과의 메타분석(또는 다른 합성) 결과에서이다.
- 포함된 연구의 내적 타당도 평가를 통해 연구결과에 비뚤림이 생길 위험, 즉 진정한 중재 효과를 과대평가하거나 과소평가할 위험을 강조해야 한다.
- 문헌고찰 저자는 연구결과의 직접성과 이질성 탐구, 연구 내 비뚤림위험성 평가, 합성에 의한 비뚤림위험 평가에 영향을 미칠 수 있는 연구자의 이해 상충을 고려해야 한다.
- RoB 2.0은 시험 설계, 실행 및 보고의 다양한 측면에 중점을 두며, 비뚤림 영역이 고정되어 있다.
- 각 영역에서 발생하는 비뚤림위험에 대한 판단은 신호 질문에 대한 대답을 기반으로 알고리즘에 의해 이루어진다. 판단은 비뚤림위험의 '낮음' 또는 '높음'일 수 있고, '일부 우려' 일 수 있다.
- 결과에 대한 전반적 비뚤림위험은 비뚤림 영역을 통틀어 가장 낮은 평가에 기초한다. 제안 된 영역 수준과 전반적 비뚤림위험 판정 모두는 재정의될 수 있지만, 정당화가 필요하다.
- RoB 2.0에 포함된 비뚤림의 영역은 무작위배정과정에서 생기는 비뚤림, 의도한 중재에서 이탈로 인한 비뚤림, 중재결과 자료의 결측으로 인한 비뚤림, 중재결과 측정의 비뚤림, 보고된 연구결과 선택의 비뚤림 등 5가지이다.

6.1 비뚤림위험 평가의 개념

6.1.1 서론

비뚤림(bias)은 연구결과에 체계적인 오류가 있는 것이다. 비뚤림으로 연구결과는 과소평가 또는 과대평가 될 수 있다. 과거 비뚤림위험은 개별 연구에서 평가되었지만 현재는 연구결과의 비뚤림과 합성 결과의 비뚤림을 모두 포함하는 개념으로 생각한다. 즉 비뚤림이 발생하는 두 장소를 구별하는데 첫 번째는 체계적 문헌고찰에 포함된 개별 연구의 연구결과이며 다른 하나는 연구결과의 메타분석(또는 다른 합성)이다. 문헌고찰의 결론은 포함된 연구결과에 따라 달라지므로 비뚤림이 있는 연구에 대한 메타분석은 잘못된 결론을 도출한다. 따라서 문헌고찰 결과를 해석할 때 연구결과의 비뚤림위험을 체계적으로 고려해야 한다. 또한 합성에 포

함되어야하는 연구를 포함하지 못해서 비뚤림이 발생할 수 있다. 특히, 적격 연구의 연구결과를 보고하는 방법, 시점 및 보고 장소에 대한 결정이 연구결과의 성격 및 방향에 영향을 받는 경우 체계적 문헌고찰의 결론이 손상 될 수 있다. 이것은 '비보고 비뚤림'이라고 한다.

비뚤림이 특정 연구결과에 어느 정도 영향을 미쳤는지를 아는 것은 대개 불가능하기 때문에 연구결과에 비뚤림이 있다고 주장하기보다는 비뚤림위험에 대해 고려하는 것이 더 적절하다.

비뚤림을 비정밀과 혼동해서는 안 된다. 비뚤림은 체계적 오류를 말하며, 동일한 연구를 여러 번 반복해도 평균적으로 잘못된 대답에 도달한다는 것을 의미한다. 비정밀은 무작위 오류를 말하며, 동일한 연구를 여러 번 반복하면 평균적으로 올바른 답을 제공한다. 정밀도는 연구 대상자 수와 이분형 중재결과의 경우 사건 수에 따라 달라지며 각 연구의 효과추정치 주위의 신뢰 구간에 반영된다. 비뚤림은 또한 연구의 외적타당도, 즉 연구결과가 다른 인구 및 환경 설정으로 일반화 될 수 있는 정도와 혼동되어서는 안 된다. 예를 들어, 특정 임상 상태를 일반적으로 경험하는 인구집단을 대표하지 않는 연구대상자들을 연구에 등록시킬 수 있다. 이 연구결과는 광범위한 인구에 일반화 가능성이 제한 될 수 있지만 그것이 기반이 되는 매우 특수한 집단에서 효과의 비뚤림이 크다고 할 수는 없다.

6.1.2 비뚤림의 경험적 근거

많은 경험적 연구에 의하면 연구의 설계, 수행 및 보고의 방법론적 특성에 의해 중재 효과 추정에 비뚤림을 초래한다. 이 근거는 주로 무작위배정 임상시험에서 보고 된 방법론적 특성과 중재 효과 추정치 사이의 연관성을 조사하는 메타 역학 연구를 기반으로 한다.

메타 역학 연구의 증거에 따르면, 눈가림과 배정 은폐가 부적절하다는 것과 같은 방법론적 특성의 영향은 중재결과의 유형에 따라 다르다. 예를 들어, 결과가 주관적으로 측정(예: 통증) 될 때 과대평가의 위험성이 커지고, 받은 중재가 무엇인지에 대한 지식에 영향을 받지 않거나, 객관적으로 측정되어 중재결과에 영향을 미치지 않을 때(예: 사망) 과대평가 위험성은 낮아진다.

메타 역학 연구의 결과에 의하면 비뚤림의 원인으로 간주되지 않는 임상시험의 특징들도 있다. 첫째 연구의 등록 여부이다. 최근의 메타 역학 연구에 따르면 등록되지 않았거나 소급 등록 된 임상 시험에 비해 전향적 등록이 이루어진 임상 시험에서 효과 추정치가 더 낮았다. 둘째, 연구의 규모이다. 규모가 작은 임상시험이 큰 연구에 비해 효과의 크기가 더 컸다. 셋째, 다기관 연구 여부이다. 단일 센터 임상시험은 다기관 연구에 비해 표본 크기를 보정 한 후에도 효과의 크기가 더 컸다.

6.1.3 비뚤림위험 평가의 절차

6.1.3.1 평가 자료원

비뚤림위험 평가를 위한 정보원은 일차적으로 출판된 논문이다. 추가적으로 임상시험 등록부 자료도 평가에 도움이 될 수 있다. 여기에서 더 나아가 임상 연구 보고서(임상 시험 방법 및

결과에 대한 자세한 내용을 제공하는 제약 회사가 준비한 문서) 및 규제 문서 등도 비뚤림위험 평가를 위한 자료원이 될 수 있다.

6.1.3.2 평가 수행

비뚤림위험 평가는 최소한 두 사람이 독립적으로 수행해야 한다. 이런 절차는 평가의 오류를 최소화하고 판단이 한 명의 선입관에 영향을 받지 않도록 하는데 필수적이다. 또한 불일치 해결 절차도 미리 정의해야 한다.

비뚤림위험 평가자는 임상적 내용 또는 방법론적 전문 기술(또는 둘 모두)을 가져야하며 비뚤림위험 도구에서 다루고 있는 여러 가지 방법론적 문제를 적절하게 이해해야 한다. 표준화된 훈련이 비뚤림 평가의 신뢰성을 크게 향상시킬 수 있다. 검토 팀은 몇몇 논문 샘플에 대해 비뚤림위험을 평가해 보고 이견이 생기는 부분을 논의하는 예비평가 작업을 고려할 수 있다. 이러한 과정은 기준이 일관되게 적용되게 하고 합의에 더 쉽게 도달 할 수 있도록 도와주는 역할을 한다. 3~6 개의 논문에 대해 시행하는 것을 추천한다.

비뚤림위험 판단에 도달하는 프로세스는 투명해야 한다. 즉 독자는 비뚤림위험이 낮음(혹은 높음)으로 평가되었다면, 왜 그렇게 평가되었는지를 알 수 있어야 한다. 이는 비뚤림위험 표(risk-of-bias table)를 제공함을 통해 해결되는데 비뚤림위험 표에는 각 영역 내에서 각 신호 질문에 대한 응답과 각 판단을 뒷받침하는 프리 텍스트가 있다.

임상시험 보고서에는 많은 중재결과들이 있기 때문에 이중 어떤 중재결과에 대해 비뚤림위험을 평가할지는 생각해야 한다. 이 경우 결과 요약표(Summary of findings table)에 포함될 수 있는 중재결과에 대해서 평가하는 것을 추천한다. 이런 표에는 대개 7개 이하의 주요 중재결과가 포함되어 있다.

6.1.4 비뚤림위험 평가 제시

메타분석에 포함된 각 연구결과와 함께 비뚤림위험 판단을 동시에 제시하는 숲 그림을 사용할 것을 권장한다(그림 6-1). 이렇게 하면 각 연구에 주어진 가중치와 함께 고려할 때, 비뚤림위험 수준에 따라 연구의 상대적인 기여도를 알 수 있게 된다. 혹은 비뚤림위험 판단(낮음, 높음. 일부우려)에 따른 연구의 상대적인 기여도를 보여주는 막대그래프를 생성할 수 있다. 이러한 목적으로 막대그래프를 3 개의 영역으로 나누는 경우 연구의 수보다는 각 범주의 연구들의 통계 정보(예: 정밀도 또는 메타분석의 가중치)에 따라 영역을 결정하는 것이 바람직하다.

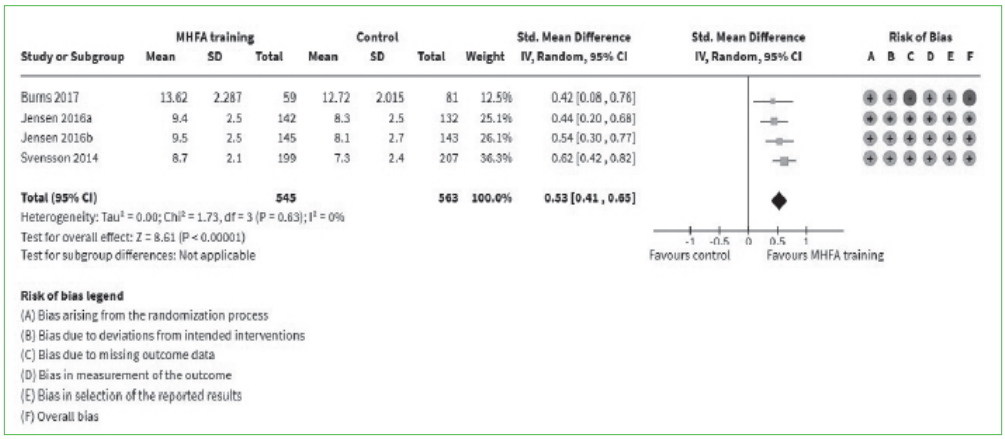


그림 6-1. 숲그림으로 비뚤림위험 제시의 예

6.1.5 비뚤림위험에 대한 요약 평가

한 연구에서 혹은 여러 연구를 통틀어 중요한 연구결과에 대한 비뚤림위험에 대해 명시적으로 요약 판단해야 한다. 코크란이 권장하는 도구(RoB 2.0 및 RoBINS-I)에서 비뚤림위험을 평가하는 경우 비뚤림위험에 대한 전반적인 판단을 하게 된다. 이러한 전반적인 판단은 개별 비뚤림 영역 평가에서 도출된다.

연구결과에 대한 비뚤림위험을 종합하여 요약하기 위해, 문헌고찰 저자는 근거 수준 평가(예: GRADE 사용) 지침을 따라야 한다. 메타분석이 비뚤림위험이 높은 연구결과에 의해 좌우될 때, 비뚤림위험이 높은 연구가 메타분석에서 제외 된 경우보다 근거 수준이 낮을 수 있다.

6.1.6 비뚤림위험 평가를 분석에 통합

일반적으로, 비뚤림위험이 높은 연구는 비뚤림위험이 낮은 연구에 비해 메타분석에서 가중치를 적게 주는 것이 합리적이다. 그러나 비뚤림위험에 따라 가중치를 적용하는 방법론적 접근법은 아직 충분히 개발되지 않았다. 다음과 같은 네 가지 방법이 있다.

첫째, 비뚤림위험이 낮은 연구에 한해 일차 분석을 시행하는 것이다. 이러한 방식으로 일차 분석을 제한하는 경우 비뚤림위험이 높은 연구가 포함된 경우 결론이 영향을 받을 수 있는지 검토하기 위해 민감도 분석을 수행하는 것이 좋다.

둘째는 다중(계층화 된) 분석을 제시하는 것이다. 전반적 비뚤림위험에 따라 계층화하면 중재 효과에 대한 효과 추정치가 도출 된다(예: 모든 연구를 기반으로 한 결과, 비뚤림위험이 낮은 연구를 기반으로 한 결과, 비뚤림위험이 높은 연구를 기반으로 한 결과). 그러나 이러한 방식으로 결과를 제시하는 것은 독자에게 혼동을 줄 수 있다. 특히, 의사결정이 필요한 경우 대개 하나의 효과 추정치를 요구하는 것이 일반적이다. 반면에 층화 된 숲 그림은 모든 정보를

투명하게 보여준다. 일반적으로 전반적 비뚤림위험을 토대로 계층화를 수행하는 것이 좋지만, 특정 비뚤림 영역(예: 무작위배정 비뚤림위험)을 기반으로 하위 그룹 분석을 수행 할 수도 있다. 전략 1과 2 중 어떤 것을 선택할지는 비뚤림위험이 높거나 불명확 한 연구가 배제 될 때 비뚤림위험은 어떻게 변하는지 정밀도는 어떻게 변하는지에 따라 저울질 할 수 있다.

셋째는 모든 연구 자료를 제시하고 비뚤림위험에 대해 기술하는 것이다. 이것은 모든 연구가 동일한 비뚤림위험을 가질 때 가능한 유일한 방법이지만, 연구들이 다른 비뚤림위험을 가지는 경우에는 두 가지 이유로 이러한 접근방식을 권장하지 않는다. 그 이유는 첫째 의사 결정이 손상된 근거에 의존할 수 있고, 둘째, 비뚤림위험이 높은 경우 가중치를 낮추지 못하여 잠재적으로 비뚤림위험이 높아질 뿐 아니라, 전체적인 중재 효과가 너무 정밀해질 수 있다. 이 방법을 이용하여 비뚤림위험 평가를 분석에 통합할 때는 GRADE 시스템을 사용하여 각 중요한 결과에 대한 근거 수준에 통합되어야한다.

넷째는 비뚤림에 대한 효과 추정을 보정하는 것으로 일반적으로 비뚤림의 크기와 그 불확실성에 대한 가정이 사전 분포를 통해 표현되는 베이지안 틀 내에서 수행 된다.

6.1.7 포함된 연구 저자들의 이해 상충 고려

이해상충이란 “일차적 이익을 위해 이루어져야 할 전문적인 판단이나 행동이 부차적인 이익에 부당하게 영향을 받을 위험이 생기는 일련의 상황”이다. 이해상충은 재정적일수도 있고 비재정적일 수도 있다.

임상 시험 설계를 할 때 핵심적인 결정이라고 할 수 있는 것은 연구대상자는 누구인지, 실험 중재는 무엇인지, 비교 중재는 무엇인지, 중재결과 평가의 시점은 언제이고 어떻게 분석할 것인지를 선택하는 것이다. 이러한 결정을 할 때는 임상적으로 그리고 과학적으로 이상과 실제 사이에서 타협이 불가피하다. 그러나 연구자에게 중요한 이해 상충이 있는 경우 임상 적용 가능성을 희생시키면서 긍정적인 임상 시험 결과가 될 가능성을 높이는 방향으로 의사결정을 할 수 있다. 임상시험 디자인 선택은 더 미묘할 수 있다. 예를 들어, 더 나은 대체 의약품이 존재하지만 열등 의약품과 비교하거나, 유해 관련 임상시험에서 고용량이 초점이지만, 저용량을 사용해서 유해 효과가 덜 나타나게 하는 경우 등이다.

코크란 비뚤림위험평가 도구 1.x 버전을 사용 할 때는 ‘다른 비뚤림의 요소’ 중 하나로 이해상충을 넣기도 하였다. 하지만 RoB 2.0에서는 비뚤림위험에 직접적으로 이해상충을 포함시키지 않는다. 대신 이해상충이 영향을 미치는 여러 가지 경로에 대해 고려하도록 권고하고 있다. 이해상충은 다음과 같은 여러 가지 경로를 거쳐 효과 추정치에 비뚤림을 초래할 수 있다.

- 연구의 직접성(예: 허약한 환자 배제, 단기 대리 결과에 집중)
- 적격 연구들의 설계상의 차이로 인한 결과의 이질성(예: 더 나은 대체 의약품이 존재하지만 열등 의약품과 비교)

- 포함된 연구결과에 비뚤림이 생길 위험(예: 선택적 보고, 무작위배정 후 배제)
- 연구결과 결측으로 인한 합성과정에서의 비뚤림위험(예: 연구 전체 미보고).

이해상충은 임상시험에서 여러 가지 메커니즘을 통해 효과 추정치에 비뚤림을 초래한다. 예를 들어, 연구대상자를 임상시험에 모집하는 사람들이 강한 이해상충을 갖고 있고, 배정 순서 은폐가 이루어지지 않으면 실험 중재에 유리하게 하기 위해 배정 과정을 뒤엎을 가능성이 있다. 마찬가지로, 강한 이해 상충을 가진 연구자는 실험 중재에 예상대로 반응하지 않은 일부 환자를 분석에서 제외하여 중재결과 자료 결측으로 인한 비뚤림을 초래할 수 있다. 또한 유리한 결과에 대한 선택적 보고는 이해 상충과 밀접하게 관련 될 수 있다.

임상시험에서 이해 상충을 평가할 때 대부분 연구논문에 선언 된 내용에 주로 의존한다. 자금 출처 관련 정보는 대부분 논문에 있지만 이해 상충은 ICMJE(International Journal of Medical Journal Committee) 선언과 같은 보조문서 형태로 부록에 있는 경우도 있다. 임상시험 출판물 및 저자 선언(예: ICMJE 선언 양식)에 보고된 정보를 기반으로 주저자 혹은 교신저자의 이해 상충을 면밀히 조사하는 것이 좋다. 이해상충이 의심되면 예상되는 작업 부하의 균형을 고려하면서 이해상충에 대한 추가 정보 소스를 검색할 수 있다. 경험적으로 자금 출처가 불분명하고, 주요 저자나 교신 저자의 이해상충 보고가 없는 경우 Open Payments 데이터베이스, ClinicalTrials.gov를 찾아보거나 주저자의 이전 출판물에서 이해상충 공개 내용을 검토 할 수 있다.

'포함된 연구의 특성'표에서 모든 연구에 대한 기금 정보 및 저자의 이해 상충 문제를 제시하는 것이 좋다. 또한 이해상충이 있는 임상 시험이 이해상충이 없는 임상 시험과 다른 중재효과 평가 또는 다른 변수를 포함하는지 하위 그룹 분석을 통해 탐색할 수 있다. 이때 문헌고찰 저자는 어느 정도의 이해 상충을 중요한 것으로 간주해야 할지에 대한 역치를 가져야 한다. 그러한 역치를 너무 낮게 책정하면 사소한 이해상충 때문에 중요한 이해상충을 흐릴 수도 있다. 너무 높게 설정하면 중요한 이해 상충이 낮게 평가되거나 무시될 위험도 있다.

6.2 무작위배정 임상시험 비뚤림위험 평가

6.2.1 서론

2008년 코크란 연합은 코크란 Risk of Bias(RoB) 도구를 발표했으며 2011년에 약간 수정되었다. RoB 도구는 다음과 같은 핵심 원칙에 따라 개발되었다.

1. 비뚤림위험이라는 단일 개념에 중점을 둔다. 보고 질, 정밀도(연구결과에 무작위 오류가 없는 정도) 또는 외적타당성(직접성, 적용 가능성 또는 일반화 가능성)과 같은 다른 개념은 고려하지 않는다.

2. 영역 기반(또는 컴포넌트) 접근 방식이며, 다양한 유형의 비뚤림을 고려한다. 무작위 순서생성, 배정순서 은폐, 연구대상자 및 연구자 눈가림, 결과평가자 눈가림, 불완전 결과 자료, 선택적 결과보고 및 기타 비뚤림의 7 가지 영역을 평가하며 점수 시스템은 없다.
3. 영역은 이론적 고려와 경험적 근거의 결합을 바탕으로 선정된다.
4. 비뚤림위험에 대한 평가는 판단을 필요로 하므로 투명해야 한다. 비뚤림의 '낮음', '높음' 또는 '불확실' 위험으로 평가된 각 영역에 대한 판단을 제공한다.

이 도구는 코크란 리뷰 및 비 코크란 리뷰 모두에 널리 사용되었지만, 실제 사용에 대한 연구결과에 의하면 일부 영역의 평가자 간 신뢰도가 낮았으며, '선택적 결과 보고'영역에 대한 이론적 배경에 대한 재고가 필요했고, '기타 비뚤림위험 영역'의 오용, 분석과 결과 해석에서 비뚤림위험 평가의 적절한 고려 결여 같은 문제들이 제기되었다.

이러한 문제를 해결하기 위해 RoB 2.0이라는 새로운 코크란 비뚤림위험 도구가 개발되었다. RoB 2.0의 주요 특징은 아래와 같다

1. 비뚤림 평가는 연구나 중재결과(outcomes) 수준보다는 개별 연구결과(results)에 대해 시행한다.
2. 비뚤림 영역의 이름을 보다 명확하게 기술하여, 다른 의미로 사용될 위험이나(예: '선택 비뚤림', '실행 비뚤림' 등), 익숙하지 않은 용어로 인해 발생하는 혼란을 줄인다.
3. 각 영역에 대한 비뚤림위험 판단에 도움이 되는 알고리즘을 제시하고, 신호 질문을 도입한다.
4. 중재 배정 효과와 중재 준수 효과를 구분하며, 이러한 개념은 '의도한 중재에서 이탈'로 인한 비뚤림위험 평가에 사용한다.
5. 분석에서 연구대상자를 제외함(예: naïve 'per protocol' analysis)으로써 발생하는 비뚤림위험에 대한 평가는 '중재결과 결측으로 인한 비뚤림'보다는 '의도한 중재에서 이탈로 인한 비뚤림'의 영역에서 하게된다.
6. '연구결과 선택적 보고' 개념을 '연구결과 선택적 비보고' 개념과 구별하며, 후자의 개념은 도구에서 제거되어 합성 수준에서 다루어진다.
7. 새 영역을 추가할 수 있는 옵션이 제거되었다.
8. 전반적인 비뚤림위험을 평가하는 명시적 과정이 도입되었다.

6.2.2 ROB 2.0의 주요 개념

6.2.2.1 연구결과기반 vs 중재결과 기반

비뚤림위험 평가를 시작하기 전에 문헌고찰 저자는 포함된 연구에서 평가할 특정 연구결과를 정할 필요가 있다. 이전 ROB 도구가 중재결과 기반 평가였다면 2.0은 연구결과 기반 평가이다.

6.2.2.2 관심효과(중재 배정 효과 vs 중재 준수 효과)

RoB 2.0 비뚤림 영역 중 하나인 '의도한 중재에서 이탈로 인한 비뚤림'에 대한 평가는 관심효과를 어떻게 정하는지에 따라 달라진다.

- 기저상태에서 중재군에 배정한 효과(의도된 대로 중재가 시행되었는지 여부에 관계없이 나타나는 효과(중재 배정 효과, intention to treat effect, 'ITT 효과'라고 함). 또는
- 시험 프로토콜에 명시된 중재를 준수하는 효과(중재 준수 효과, per protocol effect, PP 효과라고 함).

예를 들어, '특정 건강시스템에서 특정 건강 정책을 시행할지 여부'와 같은 질문은 중재 배정 효과가 적합하며(예, 선별 프로그램을 도입 할 것인지 또는 새로운 콜레스테롤 저하제를 처방할 것인가), 개별 환자에게 어떤 의사결정이 좋은가에 대한 질문은 중재 준수 효과가 적합하다(예: 특정 환자에게 선별검사를 시행할 것인가 혹은 특정 환자에게 신약을 처방할 것인가 여부), 체계적 문헌고찰을 계획할 때는 관심효과를 정의하고 그 정의 결과에 따라 적절하게 RoB 2.0을 적용해야 한다.

주요 관심효과는 체계적 문헌고찰 프로토콜에 명시되어야한다. 대부분의 체계적 문헌고찰은 중재 준수효과보다는 배정효과를 관심효과로 정하게 된다. 경우에 따라 둘 다에 관심을 기울여야 할 수도 있다.

중재 배정 효과는 모든 무작위배정 연구대상자를 포함하는 배정된 대로 분석(ITT analysis)에 의해 추정해야 한다. ITT 분석은 무작위배정의 이점을 유지하게 되어, 중재군과 대조군이 측정되거나 측정되지 않은 예후 인자에 대해 체계적으로 다르지 않게 된다.

배정된 대로 분석의 원칙은 다음과 같다.

1. 중재를 의도 한대로 받았는지 여부에 관계없이 무작위 방법으로 배정된 대로 분석한다.
2. 무작위배정에 참여한 모두를 분석에 포함 시킨다. 이것은 모든 연구대상자에 대해 중재결과 자료를 측정함으로써 이루어진다.

관심 효과가 중재배정 효과일 때, 메타분석에 포함할, 그리고 비뚤림위험을 평가할 임상시험 연구결과는 다음과 같은 우선순위에 따라 선택되어야한다.

1. 위에서 정의 된 완전 ITT 분석에 해당하는 연구결과
2. ITT 원칙을 준수하는 분석(때로는 수정 배정된대로(modified intention-to-treat', mITT) 분석이라고 설명함)에 해당하는 연구결과 : 이 경우 배제된 결측 결과를 제외하고는 ITT 원칙을 준수한다. 이 경우 결측치로 인한 비뚤림위험이 있을 수 있음).

3. 'as-treated' 혹은 naïve per-protocol 분석에 해당하는 연구결과 혹은 적격 연구대상자가 배제된 체 이루어진 된 분석.

일부는 환자들이 임상 시험 계획서대로 중재를 받았는지에 관심을 가질 수 있다(중재준수 효과). 하지만 이 때 다음과 같이 분석하면 심각하게 편향될 수 있다.

- 무작위배정이 다르게 되었다고 하더라도 연구대상자가 실제로 받은 중재에 따라 분석하는 'as-treated(치료받은 대로)' 분석.
- 중재를 준수한 연구대상자만을 분석하는 naïve 'per protocol' 분석.

6.2.2.3 주요 비뚤림위험 영역

RoB 2.0에 포함된 비뚤림 영역은 현재 무작위배정 임상시험의 결과에 영향을 미치는 것으로 알려진 모든 유형의 비뚤림을 포함한다. 여기에는 다음과 같은 비뚤림 영역이 있다.

1. 무작위배정과정에서 생기는 비뚤림(bias arising from the randomization process)
2. 의도한 중재에서 이탈로 인한 비뚤림(bias due to deviations from intended interventions)
3. 중재결과 자료의 결측으로 인한 비뚤림(bias due to missing outcome data);
4. 중재결과 측정의 비뚤림(bias in measurement of the outcome);
5. 보고된 연구결과 선택의 비뚤림(bias in selection of the reported result)

각 영역은 모두 필요하며 추가적인 영역은 없다. 표 6-1은 각 비뚤림 영역 내에서 다루어지는 이슈이다.

표 6-1. ROB2에 포함된 비뚤림 영역과 다루는 이슈

비뚤림 영역	다루어지는 이슈
무작위배정과정에서 생기는 비뚤림	• 배정순서가 무작위배정인지 여부 • 배정순서는 적절히 은폐되었는지 여부 • 중재군간 기저 상태의 차이가 있어서 무작위배정 과정의 문제를 시사하는지 여부
의도한 중재에서 이탈로 인한 비뚤림	• 연구대상자는 임상 시험 중 배정된 중재에 대해 알고 있었는지 여부 • 간병인과 중재 제공자는 임상 시험 중 배정된 중재에 대해 알고 있었는지 여부 ※ 문헌고찰 저자의 주된 관심이 배정 효과이면서 (해당되는 경우)실험적 맥락 때문에 의도한 중재에서 벗어났는지 여부(즉, 일반적인 진료환경과 다름), 그리고 만일 이런 이탈은 중재군간 유사한지 그리고 중재결과에 영향을 미칠 수 있는지 여부 중재에 대한 배정 효과를 추정하기 위해 적절한 분석이 사용되었는지 여부와 만일 그렇지 않다면 중재결과에 상당한 영향을 줄 가능성이 있는지 여부 ※ 문헌고찰 저자의 주된 관심이 배정 중재에 대한 준수효과이면서 (해당되는 경우) 중요한 비프로토콜 중재가 두 군에서 균형이 맞는지 여부 (해당되는 경우) 중재를 시행하지 않은 것이 중재결과에 영향을 미치는지 여부 (해당되는 경우) 연구대상자들이 배정된 중재를 준수했는지 여부 (해당되는 경우) 중재의 준수 효과를 추정하기 위해 적절한 분석을 시행하였는지 여부
중재결과 자료의 결측으로 인한 비뚤림	• 해당 중재결과에 대한 자료가 무작위배정된 연구대상자 전부 혹은 거의 전부에 대해 이용가능한지 여부 • (해당되는 경우) 중재결과 자료의 결측으로 인해 비뚤림이 일어나지 않는다는 근거가 있는지 여부 • (해당되는 경우) 중재결과 자료의 결측여부가 참값에 따라 달라지는지 여부(즉 두군간 결측치의 비율이 다르거나 결측의 이유가 다른지 여부)
중재결과 측정의 비뚤림	• 중재결과 측정 방법이 부적절한지 여부 • 중재결과의 측정 혹은 확인이 중재군별로 다를 수 있는지 여부 • 결과평가자가 연구대상자의 중재군이 무엇인지 알 수 있는지 여부 • (해당되는 경우) 중재군에 대한 지식에 따라 중재결과 평가에 영향을 미치는지 여부
보고된 연구결과 선택 비뚤림	• 해당 연구가 미리 결정된 계획에 따라 자료의 눈가림이 풀리기 전에 분석되었는지 여부 • 평가된 수치형 연구결과가 연구결과에 비추어 다중 중재결과 측정치 중에서 선택되었을 가능성이 높은지 여부 • 평가된 수치형 연구결과가 연구결과에 비추어 중재결과 영역에 있는 다양한 자료 분석 가운데 선택되었을 가능성이 높은지 여부

* 신호 질문에 대한 정확한 표현과 각 질문에 대한 가이드에 대한 정보를 얻으려면 www.riskofbias.info에서 전체 비뚤림 도구를 참조할 것.

6.2.3 도구의 구성

RoB 2.0 도구는 다음과 같이 구성된다.

1. '신호 질문'
2. 신호 질문 응답에 따라 판단을 매핑하는 알고리즘과 이에 따른 비뚤림위험 판단
3. 프리 텍스트 박스 : 신호 질문 및 비뚤림위험 판단에 대한 합당한 설명을 위함.
4. 비뚤림 방향을 예측(및 설명)하는 옵션.

신호질문의 목적은 비뚤림위험 평가를 체계적으로 접근하도록 하는 것이다. 대부분 사실에 근거하여 합리적으로 결정되지만, 일부는 어느 정도의 판단력을 필요로 한다. 응답 옵션은 다음과 같다.

- 그렇다(Yes, Y)
- 아마도 그렇다(Probable Yes, PY)
- 아마도 아니다(Probable No, PN)
- 아니다(NO, N)
- 정보 없음(No Information, NI)

6.2.3.1 신호질문

단순성과 명확성을 극대화하기 위해 신호 질문을 가장 자연스럽게 하고 '그렇다'의 응답에 따라 비뚤림위험이 낮거나 높을 수 있도록 하였다. '그렇다'와 '아마도 그렇다' 응답은 비뚤림위험에 대해 동일한 의미를 가지며 '아니다'와 '아마도 아니다'의 응답도 마찬가지이다. 최종 응답('그렇다'와 '아니다')은 일반적으로 신호질문과 관련하여 확고한 근거가 있음을 암시하며. "아마도'는 일반적으로 판단이 내려 졌다는 것을 의미한다.

'정보 없음' 응답은 (1) '그렇다', '아마도 그렇다', '아니다' 또는 '아마도 아니다'로 답변하기에는 세부 사항이 불충분하거나, (2) 임상시험 맥락을 감안할 때 '아마도 그렇다'또는 '아마도 아니다'라고 대답하는 것이 비합리적 때에만 사용한다. 예를 들어, 규제 임상시험이고, 경험이 많은 임상 시험 기관에 의해 시행되는 대규모 임상 시험인 경우 무작위배정 추출 방법에 대한 특정 정보가 없다고 하더라도 배정 순서 은폐에 관한 신호 질문에 대해 '정보 없음'보다는 '아마도 그렇다'라고 응답하는 것이 여전히 합리적 일 수 있다.

6.2.3.2 알고리즘

RoB 2.0 도구에는 신호 질문에 대한 응답에 따라 각 영역에 대해 비뚤림위험에 대해 판단하게 해주는 알고리즘이 포함되어 있다. 알고리즘은 신호 질문에 대한 응답을 가능한 조합해서 비뚤림위험이 낮거나 높다는 판단을 하게 해준다.

'판단'이라는 용어를 사용하는 것은 비뚤림위험 평가에 중요하다. 알고리즘은 판단을 제안하지만, 문헌고찰 저자는 이를 확인하고 적절하다고 판단되면 변경해야 한다.

6.2.3.3 프리텍스트 박스

신호 질문과 판단 옆에 있는 프리 텍스트 박스는 각 응답에 대한 근거 내용을 기록할 수 있게 하는 공간이다. 경우에 따라 동일한 정보가 두 가지 이상의 질문에 대답하는 데 사용되면 같은 내용을 여러 번 표시한다. 가능하면 연구 보고서 텍스트를 직접 인용하도록 한다. 알고리즘을 따르지 않을 경우 이유를 명기해야 한다.

6.2.3.4 전반적 비뚤림 방향

RoB 2.0에는 각 영역 및 전반적 비뚤림 방향에 대해 판단하도록 하는 부분이 있다. 일부 영역의 경우 비뚤림이 무효점으로 향하거나 멀어지는 것을 쉽게 판단할 수 있다. 예를 들어, 배정 된 중재군에서 다른 중재군으로 전환되는 비율이 높으면 중재군간 효과의 차이를 줄이는 방향으로 영향을 미쳐서 중재 준수의 추정치가 무효점 방향으로 편향되는 결과가 나타날 수 있다. 만일 문헌고찰 저자로서 비뚤림의 방향을 판단하는 명확한 근거가 없다면, 추측해서는 안 되며, 그 경우 해당 칸은 비워 둘 수 있다.

6.2.3.5 전반적 비뚤림위험

전반적 비뚤림위험의 경우도 개별 영역별 비뚤림위험과 응답 옵션이 같다. 표 6-2는 영역 내의 비뚤림위험 판단 결과에 따라 전반적인 판단을 어떻게 하는지에 대한 접근법을 보여준다.

표 6-2. 특정 결과에 대한 전반적인 비뚤림위험 평가 방법

전반적인 비뚤림위험 판단	기준
낮은 비뚤림위험	임상시험에서 현재 연구결과의 모든 영역에 대해 낮은 비뚤림위험으로 판단됨
일부 우려	임상시험에서 적어도 한 개의 영역에서 일부 우려가 있었지만 어떠한 영역도 높은 비뚤림위험이 없음
높은 비뚤림위험	임상시험에서 연구결과의 적어도 한 영역에서 높은 비뚤림위험을 보였거나, 다수의 영역에서 일부 우려사항이 있었고 이러한 우려는 연구결과의 신뢰도를 상당히 낮출 것으로 예상한다.

개별 영역에 대한 비뚤림위험 수준을 판단했다는 것은 연구결과의 전반적인 비뚤림위험이 최소한 이 정도는 된다는 의미이다. 따라서 어떤 영역이든 비뚤림위험이 '높음'이면, 평가 영역에 상관없이 높은 비뚤림위험을 갖는다고 판단해야 한다. 실제로 신호 질문에 대한 응답으로 비뚤림위험을 '높음'으로 판단한 경우, 평가자는 확인 된 문제가 전반적인 결과에 대해 이러한 판단을 보증하기에 충분한 관심사인지 여부를 고려해야 한다. 그렇지 않다면, 제시된 판단을 무효화하고, 합리적 설명을 제공하는 것이 적절한 조치일 것이다. 여러 영역에서 '일부 우려'가 있으면 결과가 높은 비뚤림위험을 갖는다고 판단할 수 있다.

6.2.4 비뚤림위험 영역

6.2.4.1 무작위배정과정에서 생기는 비뚤림

성공적으로 수행된다면 무작위배정을 통해 중재군 간의 알려진 또는 알려지지 않은 예후 인자(질병의 중증도 또는 합병증의 존재 여부와 같이 중재결과를 예측하는 요인)의 영향을 피할 수 있게 된다.

연구대상자를 무작위배정하기 위해서는 무작위배정 과정을 기반으로 배정 규칙을 정해야 한다. 이런 과정을 배정순서 생성이라고 한다. 그 다음으로 연구대상자 또는 임상시험 인력이 모집이 완료될 때 까지 배정 정보를 알 수 없도록 조치를 취해야 한다. 이를 배정 은폐라고 한다.

RoB 2.0 도구에는 기저 상태 특성을 비교해서 무작위배정 과정에 문제는 없는지 점검하는 내용이 있다. 하지만 기저상태 불균형이 있다고 해서 이것을 비뚤림위험의 증거로 해석해서는 안 된다. 불균형이 우연히 일어난 경우 체계적인 비뚤림이라고 볼 수 없기 때문이다. RoB 2.0 도구는 우연히 발생하는 기저상태 불균형은 문제 삼지 않는다.

6.2.4.2 의도한 중재에서 이탈로 인한 비뚤림

이 영역은 의도한 중재로부터 이탈되는 정도가 중재군 간에 계통적 차이가 있을 때 발생하는 비뚤림이다. 의도한 중재의 차이로 인해 발생하는 비뚤림을 실행 비뚤림(performance bias)이라고 한다. 의도한 중재는 임상시험 프로토콜에 명시된 중재이다. 이 영역에 대한 평가는 관심 효과에 따라 달라진다. 만일 중재배정 효과가 관심 효과라면, 의도한 중재로 부터의 이탈(예: 환자의 상태가 급격히 악화되어 이탈이 필요한 경우)은 비뚤림을 유발하지 않는다. 하지만 중재가 일상진료를 반영하지 않거나, 중재군간 불균형이 있거나 중재결과에 영향을 줄 때는 비뚤림을 일으킬 수 있다.

임상 시험 계획서에 명시된 중재 준수의 영향을 조사하기 위해서는 1) 중재가 얼마나 잘 수행되었는지, 2) 연구대상자가 그것을 얼마나 잘 준수했는지, 3) 의도하지 않은 비 프로토콜 중재를 의도한 중재와 함께 받았는지, 만일 그렇다면 그것이 중재군 간에 균형이 이루어 졌는지, 4) 이탈이 있다면, 적절한 통계 방법을 사용하여, 중재 준수의 효과에 대한 추정치를 계산하였는지 여부를 보아야 한다.

연구대상자, 중재 제공자가 중재 내용을 알지 못하게 하여 의도한 중재에서 이탈로 인한 비뚤림을 줄이거나 피할 수 있다('눈가림'). 눈가림에 성공하면 중재가 무엇인지를 알게 되어 발생할 수 있는 비프로토콜 중재, 오염(다른 사람이 받아야 할 중재를 받음), 중재 비준수 등을 막을 수 있다. 연구대상자, 중재 제공자에게 눈가림이 되지 않은 경우, 임상시험 결과에 영향을 주는 의도한 중재에서 이탈이 발생하여 비뚤림이 유발될 수 있다.

이 영역의 비뚤림위험은 중재결과에 따라 다를 수 있다. 예를 들어, 배정 중재에 대한 지식은 생리학적 중재결과(사망 위험 포함)에 중요한 영향을 미치지 않으며, 행동(예: 병원 방문 횟수)에는 영향을 미칠 수 있다.

중재 배정 효과에 대해서는 ITT 분석 원칙으로 분석해야 한다. 결측치가 있는 연구대상자가

제외 된 '수정 ITT'(mITT) 분석은 결측치 때문에 비뚤림이 생길 수 있다. 이는 '결측치에 의한 비뚤림' 영역에서 다루어진다. 중재 준수의 효과에 대해서는 도구 변수 접근법(Instrumental variable approach)을 이용해서 배정 중재를 받은 사람의 효과를 추정하기도 한다.

6.2.4.3 중재결과 자료의 결측으로 인한 비뚤림

중재결과 측정이 누락되면, 중재 효과 추정에 비뚤림이 생긴다. 중재결과 자료 결측의 이유는 추적관찰 소실, 연구대상자 연구 미방문, 자료 미제공, 자료 손실 등 여러 가지이다,

이 영역은 결측치를 추정하거나 설명하기 위한 절차로 인한 비뚤림을 포함하여 중재결과 자료의 결측으로 인한 비뚤림위험을 다룬다. 중재결과 자료 결측이 있는 연구대상자를 제외한 분석은 '완료 증례 분석'('complete-case' analyses)이라고 한다. 이때 만일 중재결과의 결측 여부가 참값과 관련이 없다면, 중재결과 자료 결측은 비뚤림을 초래하지 않을 것이다. 그리고 중재결과의 결측 여부가 실제 값과 관련되어 있고 중재결과 측정의 누락이 중재그룹과 중재결과의 참값에 의존한다면 중재결과의 결측은 비뚤림을 초래할 수 있다. 만일 중재결과 자료 결측이 참값과 관련이 있고, 실험 중재의 효과가 대조 중재의 효과와 다르다면 중재결과 자료 결측은 비뚤림을 유발한다.

중재결과 자료 결측의 비율에 따라 비뚤림위험을 분류하는 것은 유혹적이지만 **불행히도 중재결과 자료 결측 비율이 '충분히 작다'라고 할 수 있는 합리적인 기준은 없다.**

다음과 같은 경우는 중재결과 자료 결측에도 불구하고 비뚤림이 없다라고 확신할 수 있다.

1. 중재결과가 모든 연구대상자에서 측정된 경우
2. 중재결과 자료 결측 비율이 충분히 낮아서 어떠한 비뚤림도 중요하지 않은 경우
3. 민감도 분석(임상시험 저자 혹은 문헌고찰 저자가 실시)에 따라 중재결과 자료 결측이 중재효과에 영향이 없다고 확인된 경우

중재결과 자료 결측으로 인한 비뚤림 가능성을 해결하기 위해 통계 분석(완료증례 분석 대신 혹은 부가적으로)을 제시할 수 있다. 이러한 접근법에는 단일 결측값 대치법(single imputation, 연구대상자에게 이벤트가 발생하지 않았다고 가정 혹은 마지막 관찰이 이월되었다고 가정), 다중 결측값 대치법과 우도 기반 방법 등이 있다.

6.2.4.4 중재결과 측정의 비뚤림(bias in measurement of the outcome);

중재결과 측정 오류는 중재 효과 추정치에 비뚤림을 초래할 수 있다. 이들은 종종 측정 오류(연속형 변수의 경우), 오분류(이분형 변수 또는 범주형 변수의 경우) 또는 과소 / 과다 확인(사건형 변수의 경우)으로 불린다. 측정 오류는 중재 배정과 관련하여 차등적일 수도 있고 그렇지 않을 수도 있다.

이 영역에서 비뚤림위험은 다음 다섯 가지 고려 사항에 달려 있다.

1. 중재결과를 측정하는 방법이 적절한 지의 여부. 무작위배정 임상시험의 중재결과는 적절한

측정법을 사용하여 평가해야 한다.

2. **중재결과의 측정 또는 확인이 중재군 간에 다른지 또는 다를 수 있는지 여부.** 중재결과를 측정하거나 확인하는 데 사용되는 방법은 중재군 전체에서 동일해야 한다. 이것은 일반적으로 사전 지정된 중재결과이지만 예상치 못한 유해 효과의 경우처럼 중재결과 자료를 수동적으로 수집하면 문제가 발생할 수 있다.
3. **중재결과 평가자가 누구인지 여부.** 중재결과 평가자는
   1) 고통, 삶의 질, 자기완성 질문지와 같은 연구대상자보고 중재결과 일 때는 연구대상자이다.
   2) 중재결과가 임상진찰 결과이거나, 임상사건의 발생, 수술 중재 결정과 같은 치료적 결정인 경우는 중재 제공자이다.
   3) 판결위원회와 같이 연구대상자에게 제공되는 중재에 직접 관여하지 않는 관찰자이거나, 레지스트리에 포함 여부를 기록하는 경우는 보건의료 전문가이다.
4. **결과 평가자 눈가림이 이루어졌는지 여부.** 임상시험 중에 이중눈가림은 실현 가능하지 않은 경우에도 중재결과 평가자 눈가림은 가능하다. 그러나 연구대상자 보고결과에는 어려움이 있다.
5. **중재결과 평가가 받은 중재에 대한 지식에 의해 영향을 받을지 여부.** 결과 평가자 눈가림이 이루어지지 않은 연구에서 비뚤림위험은 중재결과 평가가 판단을 포함하는지에 따라서 달라지는데 이는 중재결과의 유형에 따라 다르다.

6.2.4.5 보고된 연구결과 선택의 비뚤림

이 영역은 연구자가 여러 중재 효과 추정치 중에서 보고하는(결과의 방향, 강도 또는 통계적 유의성을 기준으로) 연구결과를 선택하여 발생하는 비뚤림을 다룬다. 비뚤림위험에 대한 고려 시 다음을 구분해야 한다.

- 중재결과 영역(outcome domain) : 측정 방법에 관계없이 관심 있는 종료점(예: 우울증의 존재 또는 중증도)
- 특정 중재결과 측정법(예: 중재 시작 6 주 후 Hamilton 척도를 사용하여 우울 정도를 측정)
- 중재결과 분석 : 하나 이상의 중재결과 측정을 통해 얻은 구체적인 연구결과(예: 중재군간 해밀턴 등급 척도 점수의 6주 후 평균 변화)

이 영역은 연구자가 측정하고 분석 한 중재결과 영역의 선택적인 비보고(또는 불완전 보고)로 인한 비뚤림을 다루지는 않는다.

이 영역은 다음을 고려한다.

1. **눈가림이 해제된 결과 자료가 공개되기 전에 사전 지정된 최종 계획에 따라 임상시험을 분석했는지 여부.** 각 임상 시험의 사전 분석 계획을 반드시 검색하여야 한다. 그렇게 하면 결과 보고서에서 생략되었거나 추가 된 결과 측정치 또는 분석 내용을 확인할 수 있다.
2. **중재결과 영역 내에서 평가 된 여러 측정치 중에서(연구결과에 근거하여) 특정 중재결과 측정치**

만 선택적으로 보고 하는지 여부.

이러한 것의 예에는 아래와 같은 것들이 있다.

- 여러 시점 자료 중에서 하나 혹은 일부만 보고
- 여러 도구(예: 통증 척도)를 사용하였지만 가장 유리한 결과를 보이는 도구의 연구결과만 보고
- 여러 평가자가 평가하였지만(예: 의사가 평가한 것과 환자가 평가한 우울증 척도) 가장 유리한 결과를 보이는 측정치만을 보고
- 여러 하위 측정치가 있을 때 유리한 하위 측정치만을 보고

3. **중재 효과를 추정하는 여러 분석 중 특정 분석(연구결과를 기반으로)만을 선택적 보고하는지 여부.**

이러한 것의 예에는 아래와 같은 것들이 있다.

- 기저상태에 따라 보정 된 변화값과 중재 후 점수 모두를 분석하고 이 중 유리한 것만을 보고
- 예후 인자에 대한 보정을 한 경우와 하지 않은 경우를 분석하여 유리한 것만을 보고
- 연속변수를 여러 개의 분절점을 기반으로 범주형 자료로 변환
- 단일 또는 하위 집합에 대한 전체보고와 함께 여러 복합 결과에 대해 생성된 효과 추정

6.3 비뚤림위험 평가 관련 MECIR

C52: 비뚤림 위험 평가(필수)	
개별 연구에 대해 적어도 하나의 구체적인 연구결과에 대해 비뚤림위험을 평가한다. 무작위배정 임상시험은 RoB 2.0 도구를 이용하여야 하며 핸드북에 기술된 대로 해당 영역에 따라 판단 내용과 그러한 판단의 근거를 제시하여야 한다.	모든 포함된 연구에 대해 적어도 하나의 구체적인 중재결과에 비뚤림위험은 연구결과에 따라 비뚤림위험이 다를 수 있다는 점을 고려하여 명시 적으로 고려되어야 한다. 무작위 연구의 비뚤림위험은 RoB 2.0 도구를 이용하는 것이 권고된다. 하지만 다른 도구를 사용할 수도 있다.
C57: 비뚤림위험 평가 요약(권고)	
각 연구의 주요 중재결과에 대해 비뚤림위험을 요약한다.	이를 통해 연구 디자인의 특성과 그것이 연구결과에 미치는 영향 사이의 연결을 강화시키며 근거 수준을 평가하는 GRADE 접근법의 주요한 선결조건이 된다.
C58: 합성과정에서 비뚤림위험 평가에 대해 다룸(권고)	
합성과정에서 비뚤림위험에 대해 다루도록 한다(양적 합성이건 질적 합성이건). 예를 들어 전반적인 비뚤림위험에 따라 분석을 시행하거나 낮은 비뚤림위험에 대해서만 결과를 제시한다.	문헌고찰 저자는 연구의 비뚤림이 어느 정도 결과에 영향을 미치는지 고려하여야 한다. 이것은 결과의 강도를 결정하고 추후 연구가 필요한지 여부를 판단할 때 유용하다.

C59: 비뚤림위험 평가의 포함(필수)	
만일 무작위배정 비교 임상시험을 RoB 2.0과 함께 다른 도구를 이용하여 평가하였다면 주요 분석의 선택, 결론 도출을 위한 비뚤림위험 평가의 일차 도구로 RoB 2.0을 사용하여야 한다.	코크란 리뷰의 일관성을 위해서 무작위배정 비교임상시험의 비뚤림위험 평가에 두 가지 이상의 도구를 사용하였다면 RoB 2.0 도구를 더 우선시하여야 한다. RoB 2.0 도구 평가 결과는 근거수준 평가를 위해서 GRADE 접근법에 곧바로 투입될 수 있다.
C60: 포함 연구의 이해상충을 다룸(권고)	
포함 연구의 이해상충을 다루어 이들이 a) 연구 디자인의 차이, b) 연구결과의 비뚤림위험, c) 연구 합성의 비뚤림위험에 미치는 영향을 고려한다.	문헌고찰 저자는 포함 연구에 주목할만한 이해상충이 있는지를 평가하는 것을 고려한다. 이런 평가를 통해 연구의 이질성을 탐구하는데 도움을 줄 수 있고(하위분석 분석), 이해상충이 연구결과와 합성 결과를 어떻게 편향시키는지에 고려하게 한다. 이해상충에 대한 정보는 '포함연구의 특성'표에 제시할 수 있다.

6.4 참고문헌과 읽을거리

- Boutron I, Page MJ, Higgins JPT, Altman DG, Lundh A, Hróbjartsson A. Chapter 7: Considering bias and conflicts of interest among the included studies. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA(editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.0(updated July 2019). Cochrane, 2019. Available from www.training.cochrane.org/handbook.
- Higgins JPT, Savović J, Page MJ, Elbers RG, Sterne JAC. Chapter 8: Assessing risk of bias in a randomized trial. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA(editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.0(updated July 2019). Cochrane, 2019. Available from www.training.cochrane.org/handbook.

제7장 비무작위 연구 포함 결정과 비뚤림위험 평가

핵심 포인트

- 일부 체계적 문헌고찰 질문의 경우 무작위배정 임상시험만으로 대답할 수 없으며, 이런 경우 비무작위 연구를 포함시키는 것이 정당화될 수 있다.
- 비무작위 연구는 중재 효과를 평가할 때 무작위배정 임상시험에 비해 잠재적 비뚤림이 더 클 수 있으므로, 문헌고찰과 메타분석에 비무작위 연구가 포함되었을 때는 항상 신중하게 결과를 해석해야 한다.
- 중재 비무작위 연구는 인과효과를 추정하는 능력에 변이성이 있다. 연구의 핵심 설계 특성을 통해 '강한' 연구와 '약한' 연구를 구별할 수 있다.
- 포함 연구의 적격 기준, 자료 수집 및 평가는 논문에 쓰여 있는 연구설계의 라벨(예: 환자-대조군 대 코호트)보다는 연구설계의 특정 특성(예: 연구의 어떤 부분이 사전 설계되었는지)을 강조할 것을 권고한다.
- 문헌고찰 저자는 잠재적 교란 요인과 연구에 따라 달라지는 잔여 교란요인과 기타 비뚤림으로부터 초래되는 이질성을 메타분석에서 어떻게 다룰지를 고려해야 한다.
- 포함된 중재에 대한 비무작위 연구의 비뚤림위험을 평가할 때 The Risk Of Bias In Non-randomized Studies of Interventions(ROBINS-I) 사용을 권장한다.
- 문헌고찰 저자는 프로토콜에 중요한 교란 영역과 관심 공동중재를 명시해야 한다.
- ROBINS-I 평가를 시작할 때, 문헌고찰 저자는 '목표 임상시험'(target trial)을 설정해야 한다. '목표 임상시험'은 평가하려는 연구와 중재와 대상군이 같고, 비뚤림위험이 없는 가상의 실용 무작위배정 임상시험이다.
- ROBINS-I는 RoB 2.0 도구와 많은 특징을 공유한다. 특정 연구결과에 초점을 맞춘 점, 비뚤림 영역이 고정되어 있는 점, 비뚤림위험 판단을 도와주는 신호 질문이 있는 점, 전반적 비뚤림위험 판단을 하는 점 등이 그러한 특성이다.
- 신호 질문에 대한 답변에 근거하여 각 비뚤림 영역에 대한 판단 및 전반적 비뚤림위험에 대한 판단은 '낮음', '중등도', '높음(serious)' 매우 높음(critical)으로 한다.

7.1 체계적 문헌고찰에 비무작위 연구를 포함

NRSI(Non-Randomized Study for Intervention)는 중재의 효과(손해 또는 이득)를 추정하는 정량적 연구로 개인 또는 개인의 군집을 중재군에 배정할 때 무작위법을 사용하지 않는

연구로 정의된다. 그러한 연구에는 통상적인 치료 결정의 과정이나 연구대상자의 선택에 따라 배정이 일어나는 연구, 즉 흔히 관찰연구라 불리는 것이 포함된다.

중재 문헌고찰에서 특정 노출과 관련된 인과성 연구 질문과 효과성 연구 질문을 신중하게 구별할 필요가 있다. 예를 들어 영양학자들은 하루에 최소 다섯 조각 과일이나 야채를 포함하는 식단(five-a-day)이 건강 관련 효과에 어떠한 영향을 미치는지 관심을 가질 수 있고 이는 인과성 질문이다. 반면에, 공중 보건 전문가들은 five-a-day를 시행하도록 식습관의 변화를 촉진하기 위한 중재의 효과에 관심을 가질 수 있으며 이는 효과성 연구 질문이다.

7.1.1 중재 비무작위 연구를 고려하는 이유는?

대체로 NRSI를 체계적 문헌고찰에 포함시키는 것을 정당화 할 수 있는 두 가지 이유가 있다.

a) 무작위배정 임상시험으로 연구 가능한 영역이기는 하지만 이용 가능한 무작위배정 임상시험에는 간접적, 불완전한 근거만이 있는 경우. 예를 들어 장기 중재결과나 드문 중재결과에 대한 문헌고찰이 필요하지만 RCT에는 간접적인 근거만 있는 경우
b) 무작위배정할 수 없거나 무작위 임상 시험에서 연구될 가능성이 극히 낮은 중재의 효과(이득 또는 유해)의 근거를 제공. 인구집단 수준 중재(예: 법제화의 효과), 연구대상자가 강한 선호도를 가질 수 있는 중재 등

여러 비무작위 연구 체계적 문헌고찰에서 포함 이유로 제시하기는 하지만 코크란 리뷰의 맥락에서 강력한 이유는 아닌 경우는 다음과 같다.

c) 이용 가능한 NRSI의 약점에 대한 명시적 평가를 통해 RCT가 가능한지 조사한다. 예를 들어 관련 하위집단의 확인을 통해 추후 무작위배정 임상시험 설계에 정보를 제공하는 데 유용할 수 있다.
d) 중재 효과가 매우 큰 경우.
e) 무작위배정 임상시험 연구 가능한 주제이지만 소수의 무작위배정 임상시험만 이용 가능하여 중재의 효과(이득 또는 유해) 추정에 대한 추가적 근거를 제공.

이중 이유(d)의 경우 효과가 크다는 것은 암묵적으로 연구결과 중심(a result-driven) 또는 사후 주장(post hoc argument)이라는 비판이 있다.

이유(e)의 적절성은 문헌고찰이 실제로 어떻게 사용될지에 대한 기대에 크게 좌우되는데 체계적 문헌고찰의 목적이 신뢰할 수 있는 근거(무작위배정 임상시험)를 확인하는 것이라면 RCT가 없을 때 '빈 문헌고찰'(empty review)로 출판할 수 있다. 그러나 목적이 임상 및 정책 결정에 정보를 주기 위해 '가장 높은 계층'의 근거보다는 '최고의 이용 가능한' 근거를 도출할 필요가 있으면 용인될 수도 있다.

NRSI를 포함하는 문헌고찰 수행에 필요한 방법론 선택은 복잡하고 문헌고찰 결과에 영향을

미칠 수 있기 때문에, 그러한 문헌고찰을 위한 프로토콜은 훨씬 더 중요하다. NRSI를 포함하는 이유를 프로토콜에 문서화해야 한다. 또한 사용할 방법과 계획된 분석에 대한 주요 방법론적 결정의 세부적인 내용이 포함되어야 한다.

NRSI를 하나 이상의 체계적 문헌고찰 질문에 포함하기로 결정할 때마다, 문헌고찰 프로토콜에 NRSI를 검토하는 적절한 방법을 지정해야 한다. 문헌고찰이 무작위배정 임상시험과 NRSI를 모두 포함하려는 경우, 프로토콜은 두 가지 모두에 적합한 방법을 지정해야 한다. 프로토콜은 무작위배정 임상시험에는 없는 세부사항(예: 잠재적 교란 요인, 중요한 공동 중재, NRSI의 비뚤림위험 평가 및 분석 세부사항)을 명시하여야 하며 NRSI를 포함하면 문헌고찰 과정이 더 어려워지기 때문에 표준 단계에 대한 세부사항을 좀 더 자세하게 제공할 필요가 있다(예: 적합성 기준 및 문헌 검색 전략 등)

7.1.2 중재 비무작위 연구 포함 기준 결정

문헌고찰이 다룰 특정 합성 PICO를 결정하면서 NRSI를 포함할 부분을 명시하여야 한다. 이 과정에서 필요하면 이해관계자와 협의할 수 있으며, 알고리즘(그림 7-1)을 사용하여 문헌고찰에 NRSI가 포함되어야 하는지 여부와 어떤 질문에 대해 NRSI를 포함하거나 제외해야 하는지를 정당화할 수 있다.

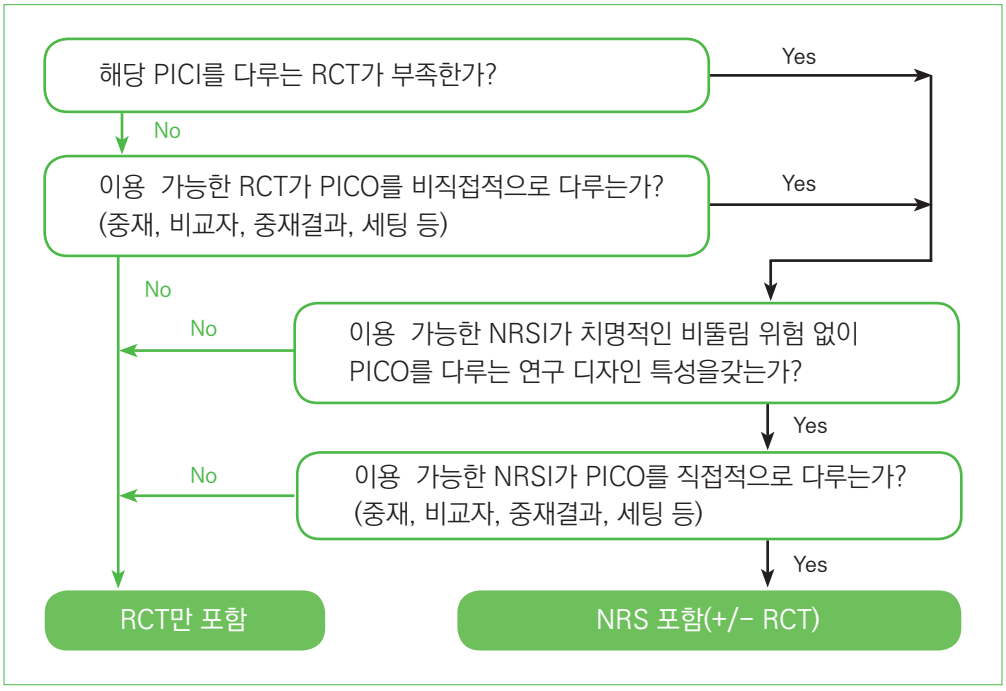


그림 7-1. 문헌고찰에 비무작위 연구가 포함되어야 하는지 여부를 결정하는 알고리즘

문헌고찰에 포함할 NRSI 유형을 결정할 때는 명시적인 연구설계 특성(연구설계 라벨이 아님)을 사용할 것을 권고한다. 이 특성은 보건 의료 중재를 평가하기 위해 가장 자주 사용되는 설계를 위한 체크리스트로 개발되었다(표 7-1).

표 7-1. 연구 특성 체크리스트. 각 항목에 대한 응답은 예, 아니오 또는 불확실로 기록한다(Reeves et al 2017).

1. 중재/비교중재는?(해당되면 둘 이상에 대해 '예'로 답할 수 있음)

개인(에게 제공/에게 투여/에 의해 선택)에게 배정
개인의 군집(에게 제공/에게 투여/에 의해 선택)에게 배정
제공된 방식으로 군집화됨(진료 의사 혹은 조직 단위에 따라)

2. 중재결과 자료가 이용가능한 시기는?(하나의 응답에만 '예'를 표시함)

중재/비교중재 이후에만(동일인에게)
중재/비교중재 이후에만(모두 동일인은 아님)
중재 전(1회) 그리고 중재/비교중재 이후에만(동일인에게)
중재 전(1회) 그리고 중재/비교중재 이후에만(모두 동일인은 아님)
중재 전 여러 번 그리고 중재 후 여러 번(동일인에게)
중재 전 여러 번 그리고 중재 후 여러 번(모두 동일인은 아님)

3. 중재 효과 추정치는 무엇인가?(하나의 응답에만 '예'를 표시함)

시간 변화에 따른 변화(동일인에 대해 다른 시점에)
시간 변화에 따른 변화(다른 시점이고 모두 동일인은 아님)
군간 차이(중재/비교 중재를 받은 개인 혹은 군집)

4. 연구자가 교란변수를 통재한 방법은?(하나의 응답에만 '예'를 표시함)

원칙적으로 모든 교란 변수를 통제할 수 있는 방법으로
원칙적으로 시간 비변동 비관찰 교란변수를 통제할 수 있는 방법으로
관찰된 공변수만을 통제할 수 있는 방법으로

5. 개인 혹은 군집 군의 형성 방법은?(해당되면 둘 이상에 대해 '예'로 답할 수 있음)

무작위배정
유사 무작위배정
연구자의 일부 다른 행동
시점의 차이에 의해
지역차이
보건의료 의사결정자 /임상진료의사
연구대상자 선호
정책결정자
중재결과에 따라
기타 방법(구체적으로 기술할 것)

6. 아래 연구의 특성은 연구가 설계된 이후에 수행되었나?(해당되면 둘 이상에 대해 '예'로 답할 수 있음)
개인/군집의 중재 전 특성
어떤 군에 이르게 하는 개인/군집의 행동/선택
중재결과 평가

7. 아래 변수들은 중재 전에 측정되었나(해당되면 둘 이상에 대해 '예'로 답할 수 있음)
잠재적 교란변수
중재결과 변수

다음과 같은 순서로 NRSI의 포함기준을 결정할 것을 권고한다.

a. 특정 PICO별로 NRSI에서 필요한 바람직한 연구설계 특성을 결정한다.
b. 범위 지정을 통해 이용 가능한 NRSI의 연구설계 특성을 표시한다.
c. 질문의 우선순위와 가능한 근거의 강도 사이의 적절한 균형을 나타내는 연구설계 특성에 근거하여 적격 기준을 설정한다.

7.1.3 중재 비무작위 연구 검색

NRSI를 검색하는 것은 무작위배정 임상시험을 검색하는 것보다 복잡하다. 일반적으로 광범위한 검색 전략을 통해 인구와 질병 특성, 중재 및 가능한 대조군에 대한 검색 문자열로 검색을 하며 이를 통해 중재에 대한 모든 근거를 확인할 수 있다.

NRSI로 검색을 제한하는 것은 어려우며, 방법론 필터를 적용하는 방법도 널리 사용되고 있지 않다. NRSI를 포함하는 체계적 문헌고찰에 시간이 많이 소요되는 특성 때문에, NRSI에 대한 검색 전략을 개발하려는 시도는 많은 수의 문헌고찰 질문을 통해 조사되지 않았다. 'snowballing' 방법을 사용하여 검색하는 것이 도움이 될 수 있는데 이 경우 하나 이상의 관련성 또는 중요성이 있는 출판물이 알려진 경우만 가능하다.

연구설계 라벨과 관련된 색인 용어로 검색하는 것을 권고하지 않는다. 오히려 NRSI 검색에 대한 전문 지식을 갖춘 정보 전문가, NRSI의 코크란 리뷰를 수행한 다른 문헌고찰 저자에게 문의해볼 수는 있다.

7.1.4 연구 선택 및 자료 수집

문헌고찰 프로토콜에 명시된 기준에 대해 포함된 연구의 적격성을 입증하기 위한 연구설계에 대한 자료. 연구설계 특성 점검표는 이를 수행하는 데 도움이 될 수 있다. 연구에서 측정한 교란 변수를 수집해야 하는데, ROBINS-I 도구는 이 정보를 수집하기 위한 템플릿을 제공한다. 또한 공동중재에 대한 정보 수집도 ROBINS-I 도구에 있는 공동 중재 정보 수집 템플릿을

이용하도록 한다. 자료수집은 조사 중인 연구 질문에 맞춰져야 할 것이다. 예상되는 다양성을 다루기 위해 시험적용을 통해 미리 적용해 보는 것을 추천한다. 자료 수집 양식을 개발하려면 여러 번 반복해야 한다.

7.1.5 비무작위연구에서 비뚤림위험 평가

7.2에서 다룬다.

7.1.6 비무작위 연구의 결과 합성

NRSI를 포함하는 문헌고찰을 작성하는 경우 비뚤림위험이 낮거나 중등도 정도이며 다른 측면에서 비교적 균질하다고 판단할 경우 메타분석을 사용하여 여러 연구결과를 결합할 수 있다. 심각한 비뚤림위험이 있는 결과를 합성하기로 결정하기는 어렵다. ROBINS-I 도구를 사용하여 비뚤림위험이 매우 높은 것으로 간주되는 연구결과는 분석에서 제외해야 한다.

서로 다른 연구설계 특성을 가진 NRSI에서 얻은 예상 중재 효과는 여러 비뚤림위험에 의해 매우 다양하게 영향을 받을 것이다. 따라서 설계 특성이 매우 다른 NRSI는 별도로 분석할 것을 권고한다. 예를 들어 무작위배정 임상시험과 NRSI를 메타분석에서 결합해서는 안 된다.

추정치 및 표준 오차에 기초한 메타분석 방법, 특히 일반적인 역분산 방법은 NRSI에 적합할 것이며, NRSI 사이의 이질성이 높을 것으로 예상되는 경우, 변동 효과 메타분석 접근법을 기본 선택해야 한다.

연구 간 이질성의 원천이 무엇인지를 조사하는 것은 비무작위 연구의 경우도 마찬가지이다. 무작위배정 임상시험보다 비무작위 연구가 더 이질적일 것으로 예상할 수 있다. 잔류 교란 요인이 연구 간 이질성의 중요한 원천이며, 교란 요인을 제어하는 데 사용되는 방법 및 교란 요인을 측정하고 분석에 포함하는 방법에도 차이가 있을 수 있다. 연구가 너무 이질적이어서 결합이 어려운 경우에는 이질성의 중요한 결정요소를 확인하기 위해 메타 회귀 분석을 수행할 수 있다. 메타분석을 실시하기 전에 주요한 연구가 결과의 결합을 정당화할 만큼 '적당한 수준'인지 검토해 보아야 한다. 연구 특성 때문에 메타분석으로 결합하기에 부적절한 경우 포함된 연구결과를 숲 그림에 표시하되 요약 추정치를 사용하지 않을 수도 있다.

7.2 비무작위 연구에서 비뚤림위험 평가

7.2.1 서론

체계적 문헌고찰에 NRSI가 포함되어 있는 경우도, 포함된 연구에 대한 비뚤림위험 평가는 필수적이다. 현재 ROBINS-I가 가장 진보적인 도구이다. ROBINS-I는 무작위배정 임상시험

비뚤림위험 평가 도구인 RoB 2.0과 많은 특성을 공유한다.

NRSI 연구결과에 대한 비뚤림위험을 평가하려면 방법론적 전문지식과 해당 질환에 대한 전문지식이 모두 필요하다. 이 과정에는 연구설계 또는 설계 특성에 대한 지식과 경험이 있는 방법론 전문가와 연구대상자 집단에 중재를 결정하는데 영향을 미치는 예측 요인에 대한 지식을 가진 중재 제공자가 참여하는 것이 권장된다. 기획 단계에서 핵심 질문을 명확하게 설정해야 하며, 교란 요인과 공동 중재를 사전에 선정하여야 한다.

NRSI에서 비뚤림은 동일한 질문을 다루고, 동일한 인구집단에 결함 없이 시행된 실용적 무작위배정 임상시험(표본 크기가 매우 큰)에서 얻은 연구결과 사이의 체계적 차이로 정의 한다. 따라서 NRSI 연구결과에 대한 비뚤림위험을 평가할 때는 각 NRSI에 대한 가상의 '목표' 무작위배정 임상시험 설정이 필요하다. 해당 무작위배정 임상시험이 실현가능하거나 윤리적일 필요는 없다. 예를 들어, 흡연의 위해를 보기 위해 무작위로 흡연군과 비흡연군으로 배정할 수 있다.

체계적 문헌고찰에 무작위배정 임상시험을 포함하는 경우, 만일 비뚤림위험이 낮다면, 해당 중재 효과는 인과적이라고 할 수 있다. 무작위배정에 의해 알려진 또는 알려지지 않은 예측 요인(질병의 심각성, 또는 동반질환의 존재와 같이 중재결과를 예측하는 요인)의 영향을 피하게 된다. 무작위배정 임상시험보다 NRSI의 비뚤림위험이 더 크다. 주요 관심사는 교란 요인, 선택 비뚤림, 정보 비뚤림에 대한 것이다.

7.2.2 주요 비뚤림 영역

7.2.2.1 교란

교란은 중재 선택과 중재결과에 공통적인 원인이 있을 때 발생한다. 교란 영역(또는 교란요인)은 개인이 관심 중재를 받을지 아니면 다른 중재를 받을지 예측하는 중재 전 예측 요인(즉, 관심 중재결과를 예측하는 변수)이다. 주요 교란 영역으로 질환 중증도, 동반질환, 보건의료 이용 정도, 의사 처방 관행, 지방량, 사회-경제적 상태 등이 있다.

조사자는 이러한 교란 영역를 특정 변수(흔히 교란 변수라고도 함)를 이용하여 측정한다. 예를 들어, 최근 체중 감소를 질환 중증도를 보정하는 데 사용할 수 있고, 이전 6개월 동안의 입원 및 병원 방문수를 보건의료 이용 정도, 지리적 특성을 의사의 처방 관행, 체질량 지수 및 허리-엉덩이 비를 지방량, 소득과 교육 상태로 사회경제적 상태를 측정할 수 있는 교란변수로 활용할 수 있다.

교란 요인은 원칙적으로 적절한 연구설계 또는 보정을 통해 극복될 수 있다. 교란 요인이 아닌 요인에 대해서는 보정을 피하는 것이 중요하다. 특히, 중재에 의해 영향을 받을 수 있는 변수(중재 후 변수)에 대한 보정은 비뚤림을 초래할 수 있다.

실제로 교란 요인을 완전히 극복하지 못하는 이유는 크게 두 가지인데. 첫째는 잔류 교란 요인이 발생하기 때문이고, 둘째는 측정되지 않은 교란 요인이 있기 때문이다.

NRSI를 문헌고찰에 포함 시킬 경우, 프로토콜 단계에서 중요한 교란 영역을 미리 지정하여

야 한다. 잠재적 교란 영역을 식별하려면 해당 주제에 대한 지식이 필요하다.

7.2.2.2 선택비뚤림

선택 비뚤림은 일부 적격 연구대상자, 추적관찰 시점 또는 일부 중재결과 사건이 제외되어 목표 임상시험에서 관찰되는 연관성과는 다른 연구결과가 나오는 것이다. 선택 비뚤림이라는 용어가 교란 요인을 의미하기도 하지만, 여기서는 교란 요인과 구별되는 의미이다. NRSI에서 선택 비뚤림이 발생하는 이유는 특정 연구대상자나 추적관찰 시점을 연구에 포함하거나('연구대상자 선정 과정에서 발생하는 비뚤림' 영역에서 다룸), 특정 연구대상자 혹은 추적관찰 시점을 연구에서 배제하였기 때문이다('자료 결측으로 인한 비뚤림' 영역에서 다룸).

선택 비뚤림은 연구대상자의 선택이나 추적관찰 시점 선택이 중재와 중재결과 모두에 관련이 있을 때 발생한다. 예를 들어, 태아 신경관 결함을 예방하기 위한 임신 중 엽산 보충에 대한 연구에서 생존 출산만 포함시키면 비뚤림이 초래될 수 있다.

선택 비뚤림은 분석에서 일부 추적관찰 시점이 제외될 때도 발생할 수 있다. 예를 들어, 신규 사용자와 비사용자를 비교하는 것이 아니라 기존 사용자(이미 중재를 받고 있는 사용자)와 비사용자를 비교할 때 비뚤림위험이 발생할 수 있다. 이것은 개시 비뚤림(inception bias) 또는 시간 지연 비뚤림(lead time bias)이라고도 불리는 선택 비뚤림의 한 유형이다.

자료 결측은 탈락, 약속 어김, 불완전한 자료 수집, 연구 책임자에 의해 일부 연구대상자가 분석에서 배제되는 등의 이유로 발생할 수 있다.

7.2.2.3 정보 비뚤림

중재 상태가 잘못 분류되거나 중재결과가 잘못 분류되거나 측정에 오류가 있는 경우 비뚤림이 발생할 수 있다. 이러한 비뚤림을 흔히 정보 비뚤림 또는 측정 비뚤림이라고 한다. 분류 오류(또는 측정 오류)는 차등적(differential)일 수도 있고, 비차등적일 수도 있다. 일반적으로 오류들이 차등적일 때 문제가 된다. 중재 상태의 차등적 오분류는 오분류 자체가 추후 중재결과나 중재결과 발생 위험과 관련이 있을 때 발생한다. 중재결과의 차등적 분류(또는 측정 오류)는 중재 상태와 관련될 때 발생한다.

중재 상태의 차등적 오분류를 방지하려면, 가능한 경우, 중재 상태를 중재결과를 모르는 상태에서 정의하고 분류해야 한다. 차등적 오분류의 잘 알려진 예는 추후 중재결과에 대한 지식 때문에 중재 분류에 영향을 미칠 수 있는 환자-대조군 연구에서 발생하는 회상 비뚤림이다.

중재결과 측정에서 발생하는 비뚤림을 결과확인 비뚤림(detection bias)이라고 한다. 이러한 비뚤림이 발생할 수 있는 상황으로는 (i) 결과 평가자가 중재 상태를 알고 있는 경우(특히 결과에 대한 평가가 주관적일 때), (ii) 서로 다른 중재군에서 중재결과 평가에 다른 방법을 사용하는(또는 관찰 강도의 차이를 두는) 경우, (iii) 측정 오차가 중재 상태(혹은 교란변수)와 관련이 있는 경우 등이다. 결과 평가자 눈가림은 중재군 간의 측정에서 체계적인 차이를 방지하기 위해 시행되지만, NRSI에서는 종종 불가능하거나 수행되지 않는다.

7.2.2.4 보고 비뚤림

NRSI에서도 무작위배정 임상시험과 동일하게 보고된 연구결과의 선택에 대한 우려가 있다. 선택적 보고는 일반적으로 연구결과가 뉴스 가치가 있거나 출판할 가치가 있을 만큼 충분히 주목할 만할 경우 발생한다. 많은 NRSI는 서면 프로토콜이 없으며, 탐구적 성격이 있어서, 설계상 중재와 중재결과 사이의 많은 연관성을 검사하게 된다.

보고된 연구결과의 선택은 중재 효과 추정의 p값, 크기 또는 방향에 기초할 경우 비뚤림으로 이어진다. 중재결과 측정의 선택에 따른 비뚤림은 복수의 측정치 중에서 특정 결과에 대한 효과 추정치를 선택할 때 발생한다.

7.2.3 ROBINS-I 도구

7.2.3.1 프로토콜 단계

ROBINS-I 평가를 계획하는 경우 프로토콜에 중요한 교란 영역을 나열해야 한다. 관련 교란 영역은 예측 요인(중재결과 예측 변수)으로 한 개인이 관심 중재 중 어떤 것을 받을지 예측하게 한다. 또한 프로토콜에 공동 중재들을 기술하여야 한다. 공동 중재는 한 개인이 관심 중재를 개시할 때 혹은 관심 중재 이후에 받을 수 있는 중재 또는 노출로, 이는 중재와 관련이 있으며 관심 중재결과에 대해 예측인자이다.

중요한 교란 요인과 공동 중재는 문헌고찰 팀의 일원인 내용 전문가의 지식을 활용하고 문헌고찰(주제 범위 문헌고찰)을 통해 확인해야 한다. 이후 가상의 목표 임상시험을 지정하여 비교될 중재 내용을 명시적으로 확인하도록 한다. 관찰연구자가 제공하는 정보가 목표 임상시험을 정의하기에 불충분하기 때문에 종종 내용에 대한 전문 지식을 필요로 한다. 다음으로 목표 임상시험에서 관심 효과를 배정 효과와 준수효과 중 어떤 것인지를 정한다. 관심 효과의 선택은 문헌고찰 저자의 결정이다.

7.2.3.2 비뚤림의 영역

ROBINS-I에 포함된 영역은 현재 NRSI 결과에 영향을 미치는 것으로 이해되는 여러 유형의 비뚤림을 다룬다. 각 영역은 필수이고, 추가 영역은 허용되지 않는다. 표 7-2에는 대부분의 NRSI 유형에 대해 적용되는 비뚤림 영역이 있다.

표 7-2. ROBINS-I 도구에 포함된 비뚤림 영역

비뚤림 영역	비뚤림의 종류	설명
중재 전 영역		
교란에 의한 비뚤림	교란	기저 상태 교란 요인은 하나 이상의 예측 변수(관심 결과를 예측하는 요인)에 의해 기저상태에서 받게 되는 중재가 예측될 때 발생한다. ROBINS-I는 시간 변동 교란 요인도 다루는데, 이것은 비교 중인 다른 중재로 전환되는 것과 기저생태 후 예측 변수에 의해 기저상태 후 받은 중재가 영향을 받는 경우 발생한다.
연구대상자 선택 비뚤림	선택 비뚤림	일부 적격 연구대상자 배제, 일부 추적관찰 시점, 또는 일부 결과사건이 배제되는 것이 중재와 중재결과 모두에 관련된 경우, 중재의 영향이 동일하더라도 중재와 중재결과 사이에 연관성이 발생한다. 이러한 유형의 비뚤림은 교란 요인과는 구별된다. 예로는 신규 사용자가 아닌 기존 사용자가 중재에 포함됨으로 발생하는 비뚤림이 있다.
중재 중 영역		
중재 분류의 비뚤림	정보비뚤림	중재 상태의 차등적 또는 비차등적 오분류에 의해 발생하는 비뚤림이다. 비차등적 오분류는 중재결과와 무관하며 일반적으로 중재의 추정 효과를 무효점으로 향하게 한다. 차등적 오분류는 중재 상태의 오분류가 중재결과 또는 중재결과 위험과 관련이 있을 때 발생한다.
중재 후 영역		
의도한 중재에서 이탈로 인한 비뚤림	교란	실험 중재군과 대조군 사이에 제공된 관리가 체계적으로 다를 때 발생하는 비뚤림이다. 이는 의도한 중재에서 이탈을 의미한다. 이 영역의 비뚤림 평가는 관심 효과(중재에 배정 효과 또는 중재 시작 및 준수의 효과)에 따라 달라진다.
자료 결측에 의한 비뚤림	선택 비뚤림	포함 연구대상자에 대한 추적관찰 결측에 의해 발생하는 비뚤림(예: 예측 요인에 의해 영향을 받는 추적관찰의 차등적 차이)이다. 혹은 중재 상태에 대한 정보나 교란 변수 등의 변수가 누락되어 발생한 비뚤림이다.
중재결과 측정의 비뚤림	정보 비뚤림	중재결과의 측정에서 차등적 또는 비차등적 오류에 의해 발생하는 비뚤림. 결과 평가자가 중재 상태를 인지하고 있거나, 다른 중재군의 결과를 평가하기 위해 다른 방법을 사용하거나, 측정 오류가 중재 상태 또는 효과와 관련이 있는 경우 발생한다.
보고된 연구결과 선택의 비뚤림	보고 비뚤림	중재결과, 분석 또는 하위집단 분석에 대한 여러 측정 중에서 결과에 따라 연구결과를 선택적 보고

7.2.3.3 ROBINS-I 도구

RoB 2.0과 마찬가지로 다음과 같이 구성되어 있다.

1. '신호 질문'
2. 영역 비뚤림위험에 대한 판단 : 알고리즘이 신호 질문 응답에 따라 판단을 도와준다
3. 신호 질문에 대한 응답과 비뚤림위험 판단을 정당화하는 프리 텍스트 상자
4. 가능한 비뚤림 방향을 예측(그리고 설명)하는 옵션

신호 질문은 영역에 대한 비뚤림위험 판단 관련 정보를 도출하는 것을 목표로 하며, RoB 2.0과 같은 방식으로 작용한다. 응답 옵션은 다음과 같다.

- 그렇다
- 아마도 그렇다
- 아마도 아니다
- 아니다
- 정보 없음

신호 전달 질문에 대한 이러한 응답에 기초하여, 영역 수준 비뚤림위험 판단의 옵션은 '낮음', '중등도', '심각함' 또는 '매우 심각함'이며, 추가 옵션으로 '정보 없음'이 있다. 비뚤림위험의 판단 내용은 RoB 2.0 도구와 다소 다르다(표 7-3).

'비뚤림위험 낮음'의 의미는 고려 중인 영역과 관련하여 잘 수행된 무작위배정 임상시험에서 비뚤림 없음에 해당한다는 점에 유의한다. 따라서 이 범주는 특히 '중재 전' 및 '중재 중' 영역에 있어 NRSI의 비뚤림위험 평가에 대한 기준점을 제공한다. 교란 요인 때문에, 비무작위 연구의 디자인 특성이나 분석의 특징 때문에 중재 효과를 연구할 때는 비뚤림위험이 낮음으로 평가되기는 어렵다. 이와는 대조적으로, 무작위배정 임상시험도 중재 후 비뚤림에는 보호 효과가 없기 때문에, 중재 후 영역에 있어서는 무작위배정 임상시험의 평가와 NRSI의 평가 사이에 더 많은 공통점이 있을 것이다. 그럼에도 불구하고 연구대상자, 보건 전문가 또는 결과 평가자의 눈가림과 같이 NRSI에서 일반적으로 가능하지 않은 무작위배정 임상시험의 다른 특징은 NRSI가 중재 후 비뚤림을 가질 위험이 더 커지게 할 수 있다.

RoB 2.0과 마찬가지로 프리 텍스트 상자는 문헌고찰 저자가 각 신호 질문 응답에 대한 판단 근거를 제시할 수 있는 공간이다. 가능한 한 간략하게, 연구 보고서의 본문에 있는 직접적인 인용문을 사용해야 한다.

도구에는 각 영역 및 전체 영역에 대한 비뚤림의 방향 판단이라는 선택적 구성요소가 포함되어 있다. 일부 영역의 경우, 비뚤림이 무효점(null)을 향하거나 무효점을 벗어나는 것으로 쉽게 판단된다.

표 7-3. 개별 비뚤림 영역에 대한 비뚤림위험 판단

비뚤림위험 판단	해석
낮음	이 연구는 해당 영역에 대해 잘 수행된 무작위배정 임상시험에 해당한다.
중등도	이 연구는 해당 영역에서 비무작위 연구로서는 건전하지만, 잘 수행된 무작위배정 임상시험에는 미치지 못한다.
심각함	이 연구는 해당 영역에서 일부 중요한 문제가 있다.
매우 심각함	이 연구는 해당 영역에서 중재 효과에 대한 어떠한 의미 있는 근거도 제시하지 못한다.
정보 없음	해당 영역에서 비뚤림위험을 판단할 정보가 없다.

모든 영역을 결합하여 도달하게 되는 연구결과에 대한 전반적인 비뚤림위험 판단에 대한 응답 옵션은 개별 영역과 동일하다. 표 7-4는 중재결과에 대한 전반적 판단에 영역 비뚤림위험 판단에 대한 매핑을 보여준다.

표 7-4. 특정 중재결과에서 전반적 비뚤림위험의 판단

비뚤림위험 판단	해석	기준
낮음	이 연구는 해당 영역에 대한 잘 수행된 무작위배정 임상시험에 해당한다.	이 연구는 해당 결과의 모든 영역에서 비뚤림위험이 낮다.
중등도	이 연구는 해당 영역에서 비무작위연구로서는 건전하지만, 잘 수행된 무작위배정 임상시험에는 미치지 못한다.	이 연구는 모든 영역에서 낮음, 혹은 중등도 비뚤림위험이다.
심각함	이 연구는 해당 영역에서 일부 중요한 문제가 있다.	이 연구는 적어도 한 개의 영역에서 심각한 비뚤림위험을 갖지만, 어떠한 영역에서도 매우 심각한 비뚤림위험을 갖지 않는다.
매우 심각함	이 연구는 해당 영역에서 중재 효과에 대한 어떠한 의미 있는 근거도 제시하지 못한다.	이 연구는 적어도 한 개의 영역에서 매우 심각한 비뚤림위험을 갖는다.

7.2.4 연구 디자인에 따른 ROBINS-I 도구

7.2.4.1 추적 관찰(코호트)연구의 비뚤림위험

ROBINS-I의 목적상, 연구 범주를 정의하게 되는데, 연구대상자를 중재 시점부터 관심 중재결과의 확인까지 관찰하는 연구를 추적관찰 연구라고 한다. 여기에는 개시 코호트 연구(연구대상자가 중재 시작 시점에서 확인됨), 비무작위배정 임상시험, 보건의료 데이터베이스 분

석, 후향적 코호트 연구 등이 포함된다.

추적관찰 연구에 대해 ROBINS-I에서 다루는 이슈는 표 7-5에 요약되어 있다. 추적관찰 연구에 대한 ROBINS-I 평가의 주요 특징은 기저상태 교란 요인(가장 익숙한 유형)과 시간 변동 교란 요인을 모두 다룬다는 것이다. 기저상태 교란 요인은 하나 이상의 중재 전 예측 요인들이 추적관찰을 시작 때 받은 중재를 예측할 때 발생한다. 일부 NRSI에서, 특히 일상적으로 수집된 자료에 기초하는 연구의 경우 연구대상자들은 시간이 지남에 따라 비교중재군으로 전환하고, 이러한 개인에서 추적관찰 시점은 어떤 시점에서 받은 중재인지에 따라 나뉜다. 기저상태 이후의 예측 변수가 연구대상자가 중재 전환에 영향을 미치는 경우, 이를 **시간변동 교란**(time-varying confounding)이라고 한다. 예를 들어, HIV 환자에 대해 각기 다른 항레트로바이러스 요법을 받은 기간으로 환자를 나누고 해당 기간 동안 중재결과를 비교한다고 해보자. 기저상태 후 CD4 셀 개수는 중재군간 이동에 영향을 미칠 수 있다. 이처럼 기저상태 이후 예측 변수가 중재 자체에 영향을 받는 경우(예를 들어, 항레트로바이러스 요법이 기저상태 이후 CD4 셀 갯수에 영향을 미칠 수 있음), 이를 **치료-교란 환류**(treatment-confounder feedback)가 있다고 말한다. 이는 전통적인 보정(예: 포아송 또는 콕스 회귀 모형)이 시간 변동 교란 요인을 제어하는 수단으로서 적절하지 않다는 것을 의미한다. 중재의 유해효과와 같은 다른 기저상태 이후 예측 요인도 중재 간의 전환을 예측할 수 있다.

표 7-5. 추적관찰 연구에 대한 ROBINS-I tool

비뚤림 영역	다루어지는 이슈
교란에 의한 비뚤림	• 기저상태 교란 요인만 고려하는지 혹은 기저상태 교란 요인과 시간 변동 교란 요인을 모두 고려해야 하는지 여부(받은 중재에 따라 추적관찰 시점이 분할되는 연구) 모든 중요한 교란 영역이 통제되었는지 여부 • 가능한 교란변수가 교란 영역을 타당하고 신뢰할 수 있게 측정했는지 여부 • 교란 요인 통제를 위해 적절한 분석 방법을 사용하였는지 여부
연구대상자 선택 비뚤림	• 연구대상자 선택이 중재 시작 후 관찰된 연구대상자 특성에 기초하였는지 여부 (해당되는 경우) 그런 특성이 중재와 관련이 있으며 중재결과(또는 중재결과의 원인)에 의해 영향을 받는지 여부 • 추적관찰의 시작과 중재의 시작은 동일하였는지 여부 • (해당되는 경우) 선택 비뚤림의 교정을 위해 보정 방법을 사용하였는지 여부
중재 분류의 비뚤림	• 모든(또는 거의 모든) 연구대상자에 대해 중재 상태가 올바르게 분류되었는지 여부 • 중재군을 분류하는 데 사용된 정보가 중재 시작 시점에 기록되었는지 여부 • 중재결과 위험 또는 중재결과에 대한 지식에 의해 중재 분류가 영향을 받는지 여부

의도한 중재에서 벗어나 발생하는 비뚤림	• 관심 효과가 배정인 경우(제7.2.3절 참조): • 임상시험적 맥락(즉, 일상 진료를 반영하지 않은 이탈) 때문에 의도한 중재에서 이탈이 있는지, 그렇다면 이탈은 군간에 균형을 이루는지, 결과에 영향을 미쳤을 가능성이 높은지 여부 • 관심효과가 중재에 대한 준수인 경우(7.2.3절 참조): • 중요한 공동 중재가 중재군 간에 균형을 이루었는지 여부 • 중재 실행을 못했을 때 중재결과에 영향을 미칠 수 있고, 중재군 간에 균형이 맞지 않는지 여부 • 연구대상자들은 배정된 중재법을 준수했는지 여부, 그렇지 않은 경우 중재군 간에 비준수가 불균형했는지 여부 • (해당되는 경우) 적절한 분석을 사용하여 중재에 대한 준수 효과를 추정했는지 여부
자료 결측에 의한 비뚤림	• 중재결과 자료 결측으로 인해 분석에서 제외된 연구대상자 수가 적은지 여부 • 중재여부에 대한 자료 결측으로 인해 분석에서 결측된 연구대상자 수가 적은지 여부 • 분석에 필요한 다른 변수 자료가 결측되어 분석에서 제외된 연구대상자 수가 적은지 여부 • (해당되는 경우) 중재결과 자료 결측으로 인한 연구결과 편향이 없다는 근거가 있는지 여부 • (해당되는 경우) 중재결과 자료 결측은 결과의 참값에 따라 달라질 수 있었는지 여부(예: 중재군간 중재결과 자료 결측 비율이 다르거나 중재결과 자료 결측 이유가 다른 경우)
중재결과 측정의 비뚤림	• 중재결과의 측정 방법이 부적절했는지 여부 • 중재결과의 측정 또는 확인이 중재군마다 다를 수 있는지 여부 • 결과 평가자는 연구대상자가 받은 중재에 대해 알고 있는지 여부 • (해당되는 경우) 중재결과 평가가 받은 중재에 대한 지식에 의해 영향을 받는지 여부, 실제로 이것이 가능한지 여부
보고된 연구결과 선택의 비뚤림	• 평가되는 수치 연구결과는 연구결과에 따라 다수의 연구결과 측정치 중 선택되었을지 여부 • 평가되는 수치 연구결과는 그 연구결과에 따라 자료의 다수의 분석 중 선택되었을지 여부 • 평가되는 수치 연구결과는 연구결과에 따라 여러 하위 그룹에서 선택되었을지 여부

7.2.4.2 비통제 전후 연구(단절적 시계열 연구 포함)의 비뚤림위험

일부 연구에서 중재결과 변수에 대한 측정이 중재가 발생하기 전과 후에 이루어진다. 연구의 분석 단위에 따라 개인, 개인 군집 또는 행정 단위(administrative entities)에 대한 분석이 수행된다. 분석에 기여하는 모든 단위가(동일한) 중재를 받고 통제가 없는 연구가 비통제 전후 연구이다.

이 연구 범주에는 단절적 시계열 연구(ITS)를 포함된다. ITS 연구는 집계 수준(aggregate level)에서 측정된 종적자료를 수집하며, 중재 실행 전 서너 번 측정 하고 중재 실행 후 서너 번 측정하게 된다. '단절'로 알려진 특정 시점은 '전'(또는 '전 중재')과 '후'(또는 '후 중재') 시점의 구분이 이루어지는 시점으로 정의한다. ITS 자료는 일반적으로 단일 시계열이며, 시계열 방법(예: ARIMA 모델)을 사용하여 분석할 수 있다. ITS 분석에서 '비교그룹'을 두며 이들은 중재(또는 간섭)를 받지 않았다. 이 때 중재 효과는 중재 후 관찰된 결과 궤적을 중재가 없었던 것으로 가정된 궤적과 비교하여 추정한다.

또한 이 범주에는 여러 개인을 중재 전과 후 각각 측정하는 연구가 포함된다. 중재 전과 후

측정이 여러 번 있을 수 있다. 이러한 연구는 통제되지 않는 종단 연구설계로 특징지어질 수 있다(대체적으로 반복적인 측정 연구, 전후 연구 혹은 pre-post studies or reflexive control studies라고도 한다). 한 가지 특별한 사례로 여러 연구대상자에 대해 한 번의 중재 전 중재결과 측정과 한 번의 사후 중재결과 측정을 하는 연구가 있다. 그러한 연구는 보통 사후 변화가 중재에 의한 것인지 여부를 판단하는 것이 불가능하기 때문에 비뚤림위험이 심각하거나 매우 심각하다고 판단될 것이다.

비통제 전후 연구의 ROBINS-I 평가에서 다루는 주요 이슈는 아래 표 7-6에 요약되어 있다. 우리는 비통제 연구가 의도한 중재를 준수하는 효과를 검토할 것으로 예상하지 않기 때문에, 중재에 대한 배정 효과만을 다룬다.

표 7-6. 비통제 전후 연구에 대한 ROBINS-I tool

비뚤림 영역	다루어지는 이슈
교란에 의한 비뚤림	• 중재 전 시점에서 충분한 중재결과 측정이 수행되어 중재 전 경향 및 패턴의 특성화가 가능한지 여부 • 중재결과에 영향을 미칠 수 있는 중재 시점과 관련된 외부 사건 또는 상황 변화 여부 • 연구자가 시간 추이 및 패턴, 주요 교란 영역에 대한 통제에 적절한 분석 방법을 사용하였는지 여부
연구대상자 선택의 비뚤림	• 추적관찰 연구와 유사. 사전 중재에서 사후 중재까지 특정 단위 그룹을 전향적으로 조사한 연구의 경우 선택 비뚤림위험이 낮다. 인구집단에 대한 반복적인 단면 조사의 경우, 연구가 전향적인 경우에도 선택 비뚤림의 가능성이 있다.
중재 분류의 비뚤림	• 중재 전 시점과 중재 후 시점 구분의 특정화가 중재결과에 영향을 미치는지 여부
의도한 중재에서 이탈로 인한 비뚤림	• 관심 효과가 중재의 배정 효과인 경우(제7.2.3절 참조): • 중재의 준비 단계(단절 전)의 효과가 충분히 고려되었는지 여부
자료 결측에 의한 비뚤림	• 개별 연구대상자뿐만 아니라 전체 군집에 대한 결과 자료가 결측되었는지 여부
중재결과 측정의 비뚤림	• 중재결과 평가 방법이 중재 전후에 비교가능한지 여부 • 중재 수행과 동시에 중재결과 측정에서 체계적 오류의 변화가 있는지 여부
보고된 연구결과 선택의 비뚤림	• 추적관찰 연구의 이슈와 동일함

7.2.4.3 비교 전후 연구에서 비뚤림위험

비교 전후 연구(controlled before-after studies, CBA)는 (i) 중재군과 대조군(비치료군 혹은 비교 중재군)으로 비무작위로 배정되며, (ii) 중재 실행 전후에 중재결과 변수의 최소 한 가지 측정이 두 그룹 모두에서 수행되는 연구를 말한다. 이 범주에는 비교 단절적 시계열 연

구(controlled interrupted time series, CITS)도 포함된다. 연구에 포함된 단위는 개인, 군집 또는 행정 단위일 수 있다. 중재는 개인 수준 혹은 일부 집계(클러스터)수준에서 이루어질 수 있다. 연구는 시간 경과에 따라 동일한 단위를 추적관찰하기도 하고(within-person 혹은 within-unit longitudinal designs이라 불리기도 함), 다른 시점에서 다른 단위에 대해 시행되기도 한다(반복적인 단면 설계라고하기도 하며 관심 인구집단이 지역적이나 보건의료기관의 상호작용을 통해 정의되고 시점에 따라 다른 사람을 포함하기도 함).

비교 전후 연구에 대한 ROBINS-I 평가에서 다루는 주요 이슈는 아래 표 7-7에 요약되어 있다.

표 7-7. 비교 전후 연구에 대한 ROBINS-I tool

비뚤림 영역	다루어지는 이슈
교란에 의한 비뚤림	• 중재 전 시점에서 충분한 중재결과 측정이 수행되어 중재 전 경향 및 패턴의 특성화 가능 여부 • 중재결과에 영향을 미칠 수 있는 중재 당시의 외적 사건이나 상황 변화가 두군 모두에서 똑같이 경험되는지 여부 • 중재 전 경향이나 중재결과의 패턴은 적절하게 분석되었으며 두 군에서 유사하였는지 여부
연구대상자 선택의 비뚤림	• 추적관찰 연구와 유사함. 인구집단에 대한 반복적인 단면 조사의 경우, 반복적인 연구대상자/단위 유형의 변화가 중재군과 비교군에서 다른 경우 선택 비뚤림의 가능성이 있다.
중재 분류의 비뚤림	• 중재 전 시점과 중재 후 시점 구분의 특정화가 중재 후 중재결과 자료에 영향을 미치는지 여부
의도한 중재에서 이탈로 인한 비뚤림	• 관심 효과가 중재의 배정 효과인 경우(제7.3.3절 참조): • 추적관찰 연구의 이슈와 동일함
자료 결측에 의한 비뚤림	• 개별 연구대상자뿐만 아니라 전체 군집(다수인의 단위)에 대한 결과 자료가 누락되었는지 여부
중재결과 측정의 비뚤림	• 중재결과 평가 방법은 중재 전후에 비교가능한지 여부 • 중재 수행과 동시에 결과 측정에서 체계적 오류의 변화가 있는지 여부
보고된 연구결과 선택의 비뚤림	• 추적관찰 연구의 이슈와 동일함

7.3 참고문헌과 읽을거리

- Reeves BC, Deeks JJ, Higgins JPT, Shea B, Tugwell P, Wells GA. Chapter 24: Including non-randomized studies on intervention effects. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA(editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.0(updated July 2019). Cochrane, 2019.

Available from www.training.cochrane.org/handbook.
- Sterne JAC, Hernáan MA, McAleenan A, Reeves BC, Higgins JPT. Chapter 25: Assessing risk of bias in a non-randomized study. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA(editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.0(updated July 2019). Cochrane, 2019. Available from www.training.cochrane.org/handbook.
- Reeves BC, Wells GA, Waddington H. Quasi-experimental study designs series-paper 5: a checklist for classifying studies evaluating the effects on health interventions-a taxonomy without labels.J Clin Epidemiol. 2017 Sep;89:30-42

제8장 비보고 비뚤림 평가

핵심 포인트

- '비보고 비뚤림'('non-reporting bias')은 적격 연구의 연구결과를 어떻게, 언제, 어디에 보고할 것인가에 대한 결정이 연구결과의 p값, 크기 또는 방향에 의해 영향을 받는 것을 말한다.
- 여러 유형의 비보고 비뚤림이 있으며 연구 보고서와 연구결과가 있을 수 있는 모든 가능한 출처를 검색해야 한다.
- 깔때기 그림(그리고 깔때기 그림 비대칭성 검사)은 문헌고찰 저자가 프로토콜이나 임상시험 등록 자료를 이용할 수 없는 경우 비보고 비뚤림의 증거를 식별하는 데 도움이 될 수 있다. 하지만, 이런 접근에는 명백한 제한점이 있다.
- 깔때기 그림 비대칭의 증거가 있는 경우, 비보고 비뚤림은 여러 가지 가능한 설명 중 하나이다. 이 때 문헌고찰 저자는 비대칭의 근원을 이해하려고 시도해야 하며, 추가적인 누락 결과를 시사하는 여러 가지 신호를 조사하고 민감도 분석 등을 시행하여 그 의미를 고려해야 한다.

8.1 개요

체계적 문헌고찰은 포함 기준을 사전에 지정하여 이를 충족하는 모든 연구를 찾아내어 특정 중재의 이득과 유해에 대한 평가를 하게 된다. 하지만 이러한 체계적 문헌고찰의 연구 목적은 특정 연구결과가 통계적 유의미성, 효과의 크기, 효과의 방향에 따라 결측될 때는 달성하기 어렵게 된다. 특정 중재에 대한 연구결과 다른 중재에 비해 효과적이며 그것이 '통계적으로 유의미' 하면 그렇지 않은 연구결과에 비해 신속하게 출판되며, 영향력이 큰 저널에 출판되고, 더 많이 인용되기 때문에 체계적인 문헌고찰 과정에서 더 쉽게 식별된다. 지금까지 이러한 형태의 비뚤림에 대해 '보고 비뚤림'이라는 용어를 사용하였지만 **'비보고 비뚤림'**이 더 적합한 용어로 보인다.

비보고 비뚤림의 주요 유형과 해당 유형의 정의는 표 8-1에 있다.

표 8-1. 다양한 비보고 비뚤림과 해당 유형의 정의

보고 비뚤림 유형	정의
출판 비뚤림	연구결과의 성격과 방향에 따라 연구결과의 출판 또는 비출판이 결정된다.
시간 지연 비뚤림	연구결과의 성격과 방향에 따라 연구결과가 신속하게 또는 지연되어 발표된다.
언어 비뚤림	특정 언어로 된 연구결과의 성격과 방향에 따라 연구결과의 출판이 결정된다.
인용 비뚤림	연구결과의 성격과 방향에 따라 연구결과가 인용되거나 또는 인용되지 않는다.
중복출판 비뚤림	연구결과의 특성 및 방향에 따라 연구결과를 여러 번 또는 한 번 출판된다.
위치 비뚤림	연구결과의 특성 및 방향에 따라 표준 데이터베이스에서 접근이 용이하거나 색인 생성 순서가 결정된다.
선택적(비)보고 비뚤림	연구결과의 성격과 방향에 따라 일부 결과 또는 분석이 선택적으로 보고한다.

가장 대표적인 비보고 비뚤림으로 출판비뚤림이 있다. 연구결과의 출판여부는 연구결과의 성격과 방향에 영향을 받는다는 설득력 있는 근거가 있다. 연구결과의 성격과 방향에 따라 연구 완료와 연구결과 발표 사이의 기간이 영향을 받을 수 있으며('시간 지연 비뚤림'), 영어가 모국어가 아닌 연구자는 주요 생의학 데이터베이스에 색인되지 않은 현지 비영어 저널에 논문이 출판될 가능성이 높다('언어 비뚤림'). 논문의 인용가능성은 연구결과의 성격과 방향에 의해 영향을 받는 것으로 보이며('인용 비뚤림'), 통계적으로 유의미한 결과를 가진 연구나 더 큰 치료 효과가 있는 논문이 여러 번 출판될 가능성이 크다('다중(중복) 출판 비뚤림'). 또한 연구결과의 특성과 방향에 따라 저널 색인에 대한 접근성 또는 수준이 차이가 날 수 있으며('위치 비뚤림'), 상당량의 정보를 적은 분량으로 압축 할 필요가 있어서 연구결과의 성격과 방향에 따라 일부 중재결과에 대한 연구결과가 결측 될 수 있거나 전혀 보고되지 않을 수도 있고('연구결과의 선택적 비보고'), 불완전한 보고도 가능하다('연구결과의 선택적 과소보고'). 연구결과의 선택적 과소보고의 예는 요약 통계나 효과 추정치를 정확하게 제공하기보다 "P〉 0.05"라고만 표시하는 것이다.

비보고 비뚤림은 체계적 문헌고찰 과정에서 연구결과 결측으로 인한 비뚤림을 유발한다. 이 경우 메타분석과 같은 근거 합성결과는 연구결과의 p값, 크기 또는 방향 때문에 일부 적격 연구결과를 사용할 수 없을 때 연구결과 결측으로 인해 비뚤림이 발생할 위험이 있다.

연구결과(results)와 중재결과(outcomes)라는 용어에 대한 구별이 필요하다. 연구결과(results)는 점 추정치(평균 차이 또는 위험 비율 등)와 그 정밀도 측정치(예: 신뢰 구간)의 조합을

설명한다. 반면 중재결과(outcomes)는 어떤 사건 발생에 대한 것이다. 완전히 정의된 중재결과는, 중재결과 영역, 특정 측정치, 특정 메트릭 및 시점이 포함된다(예: "몽고메리-아스버그 우울증 등급 총점 기준에서 8주, 50% 변화"). 이 중재결과에 상응하는 연구결과(results)는 개인자료의 집적 방법(a method of aggregation across individuals)을 추가로 요구한다. 여기서는 95% 신뢰 구간을 갖는 위험비가 될 수 있으며, 이는 중재결과를 가진 사람의 비율에서 그룹 간 차이를 추정할 수 있다.

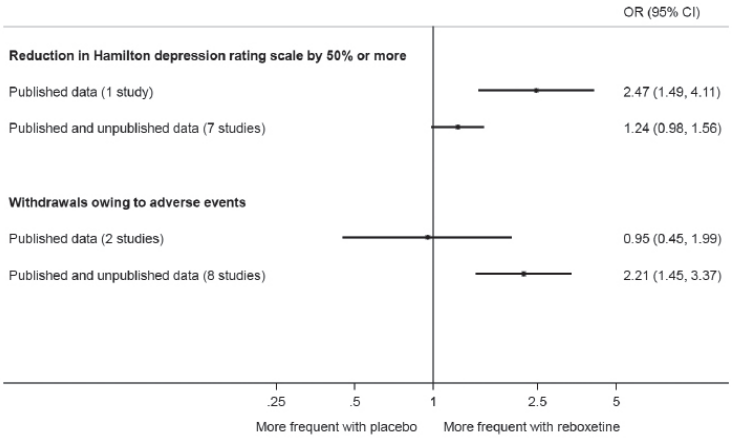
8.2 연구결과 결측으로 인한 비뚤림위험 최소화

사용된 연구결과와 결측된 연구결과는 체계적으로 다르기 때문에 합성결과에 비뚤림위험이 발생할 수 있다. 연구결과 결측으로 인한 비뚤림위험을 최소화하기 위한 방법은 여러 가지가 있지만 대표적인 것으로 1) 발표된 보고서 이외의 출처로부터의 결과 포함(임상시험 등록부의 임상 연구 보고서 및 기타 규제 문서로부터의 연구결과 포함), 2) 개시(inception) 코호트의 두 가지 활용 방법이 있다. ClinicalTrials.gov는 2008년에 등록된 임상시험에 대한 요약 연구결과를 받기 위한 자료 필드를 만들었으며 2018년 현재 약 12%에서 요약 연구결과를 확인할 수 있다. ClinicalTrials.gov와 유럽 연합 임상시험 등록부에 연구결과를 게시하기 위한 몇 가지 이니셔티브가 진행 중이므로, 이러한 등록부를 검색하는 것은 향후 체계적 문헌고찰에서 비뚤림을 최소화하는 중요한 방법이 될 수 있다.

규제 관련 중재(예: 의약품, 생물학적 제제)에 대한 문헌고찰에서 임상 연구 보고서(CSR)와 FDA 승인 패키지를 찾아보면 연구결과 결측으로 인한 비뚤림위험 최소화에 도움을 받을 수 있다. CSR 및 규제 문서는 규제 관련 중재에 대한 체계적 문헌고찰의 신뢰성을 개선할 수 있는 큰 잠재력을 가지고 있지만(Box 8-1), 이에 접근하여 그 안에 있는 자료를 얻기 위해서는 상당한 노력이 필요하다(Box 8-2)

Box 8-1. 주요 우울증에서 미출판 자료 포함 여부에 따른 reboxetine의 효과

Eyding 등 연구(Eyding, 2010)는 출판된 문헌 이외의 검색의 중요성에 대한 두드러진 예를 보여준다. 연구진은 심각한 우울증에 대해 reboxetine 대 위약 임상시험 자료와 제조자(Pfizer, Berlin)의 미발표 자료를 찾았다. 확인된 13건의 임상시험 중 26%만이 논문으로 발표되었다. 메타분석을 출판된 결과로 제한되었을 때 reboxetine는 훨씬 더 좋은 효과를 보였다(아래 그림). 예를 들어, 출판된 연구에 대한 메티분석 결과는 출판된 임상시험과 미발표 임상시험의 합친 메타분석과 비교했을 때, 우울증의 의미 있는 감소 가능성이 훨씬 더 컸다. 이와 유사하게, 두 개의 발표된 임상시험에 대한 메타분석 결과는 유해효과 때문에 철회한 환자의 수에서 두 군 간에 거의 차이가 없었지만, 6개의 미발표 임상시험이 추가되었을 때, reboxetine 환자군에서 유해효과로 인한 철회 가능성이 2배 이상 높았다.



Box 8-2. 규제 문서에 대한 정보

RIAT(Restoring Invisible and Abandoned Trials) 지원 센터 웹 사이트에는 규제 문서의 분류 체계, 관련 용어의 용어집, 규제 기관에게 CSR을 요청하는 방법에 대한 지침 및 요청을 위한 연락처 정보를 포함하여 유용한 정보를 제공한다.

8.2.1 개시 코호트(inception cohorts)

적격 연구의 유형을 제한하면 연구결과 결측으로 인한 비뚤림위험을 줄일 수 있다. 체계적 문헌고찰은 전통적으로 종결된 연구를 포괄적으로 검색하기 때문에, 비보고 비뚤림, 색인 불

량, 기타 요인 때문에 모든 연구가 실제로 식별되었는지 여부를 알 수 없게 된다. 이를 해결하기 위한 대안적 접근법으로 연구가 개시된 모든 연구(개시 코호트)에 대해 문헌고찰을 시행하는 것이 있다. 개시 코호트는 개시된 것으로 확인된 일련의 연구들을 말하며, 사전 등록된 연구에서 목록을 만든다(예: 임상시험 등록부에서만 연구 선택).

포함 연구를 등록된 연구로 제한하면 적격 연구가 결측될 가능성을 피할 수 있다. 그러나 이러한 연구에도 여전히 결측된 연구결과가 있을 가능성이 있다. 따라서 확인된 각 연구에 대한 연구결과가 존재하는지 이용가능한지를 평가해야 한다. 만일 사전 등록된 연구에서 선택적 비보고 또는 연구결과의 과소보고가 없다면 연구결과 결측으로 인한 비뚤림위험은 없을 것이다. 반대로 일부 연구결과가 선택적으로 결측된 경우, 특히 결측된 자료의 총량이 큰 경우 합성결과에 비뚤림위험이 있을 수 있다.

개시 코호트를 임상시험 등록부에만 의존하여 만드는 것이 항상 이상적인 것은 아니다. 임상시험 등록은 2004년 이후에야 증가하기 시작했는데, 그 이유는 국제 의학 저널 편집위원회 협의회(ICMJE)가 등록되지 않은 임상시험은 더 이상 출판하지 않겠다고 발표하였기 때문이다. 이러한 이유로, 이 시점 이전에 시행된 임상 시험의 경우 사전 등록된 임상시험 계획서를 찾지 못할 가능성이 크다. 또한 아주 최근까지도 규제 임상시험이 아닌 경우 사전 임상시험 등록을 하여도 그에 대한 인센티브가 거의 없었기 때문에, 사전 등록 비율이 획기적으로 증가되지 않는 한, 체계적 문헌고찰 과정에서 꽤 많은 임상시험을 확인 못할 가능성이 있다.

따라서 체계적 문헌고찰을 개시 코호트로 제한한다고 결정하는 것은 비뚤림, 정밀도, 적용가능성 사이에서 어느 정도 절충을 한다는 의미이다. 예를 들어, 사전 등록된 임상시험으로 제한하여 메타분석을 시행하면 선택적으로 연구결과가 결측되지 않은 경우 연구결과 결측으로 인한 비뚤림위험을 피할 수 있다. 그러나 소규모의 사전 등록 임상시험만 있는 경우 메타분석의 정밀도는 낮아질 수 있다. 또한 메타분석에서 얻은 요약 추정치는 사전 등록 임상시험의 핵심질문이 미등록 또는 후향적으로 등록된 임상시험의 핵심 질문보다 범위가 좁을 경우 검토 질문에 대한 적용가능성이 제한적일 수 있다. 따라서 종합 결과를 해석할 때는 정밀성과 적용가능성을 고려할 필요가 있다.

8.3 합성과정에서 연구결과 결측으로 인한 비뚤림위험 평가를 위한 틀

합성과정에서 연구결과 결측으로 인한 비뚤림위험 평가를 할 때는 다음과 같은 6단계로 접근하는 것이 바람직하다.

1. 연구결과 결측으로 인한 비뚤림위험을 평가하기 위한 합성 결과(메타분석 등)을 선택한다.
2. 해당 합성결과에 적격(eligible)인 연구결과를 정의한다.
3. 연구자가 생성한 것으로 알려진(또는 추정된) 연구결과가 이용가능하지 않아 개별합성 결과에서 결측되었는지 여부를 확인하여 기록한다 : 알려진 무지

4. 확인된 연구에서 결측된 연구결과 때문에 개별합성결과가 비뚤릴 가능성이 있는지 고려한다.
5. 개별합성 결과에서 결측 가능성이 높은 추가 연구들이 있는지 여부를 고려한다 : 알려지지 않은 무지
6. 개별합성 결과에 대한 연구결과 결측으로 인한 전반적 비뚤림위험 판단에 도달한다.

8.3.1 비뚤림위험을 평가하기 위한 합성 결과 선택

모든 합성결과보다는 환자와 보건의료 전문가에게 중요한 합성 결과에 대해서 연구결과 결측으로 인한 비뚤림위험을 평가하는 것이 좋으며 일반적으로 '결과 요약' 표에 포함되는 것이 그러한 연구결과라고 할 수 있다.

8.3.2 해당 합성결과에 적격인 연구결과를 정의

합성 결과에 포함할 수 있는 연구결과의 적격 정도는 계획된 합성 결과의 특이도에 따라 달라진다(예: 특정 시점(6주)에, 특정 측정도구(BDI)를 사용하여 우울 정도를 측정 vs 어떤 시점에서든 정신 건강 결과 정도를(예: 우울증 또는 불안)를 측정). 합성 유형이 구체적일수록 주어진 연구결과가 적격일 가능성은 낮다. 예를 들어, 2주째 BDI에 대한 결과만 있다면 이 임상시험은 최대 12주 이내의 합성결과에는 포함할 수 있지만, '6주 째 BDI 점수'에 대한 합성에는 포함될 수 없다.

8.3.3 연구자가 생성한 것으로 알려진 연구결과를 확인 후 기록(알려진 무지, 'known unknowns')

개별 합성결과에 대해 적격 연구결과가 정의되면, 확인된 모든 연구에 대해 해당 연구결과의 가용성을 조사할 수 있다. 이 때 고려해야 할 핵심 질문에는 다음과 같은 것이 있다.

1. 어떤 연구에서든 내가 찾고 있는 특별한 연구결과를 이용할 수 없는 경우가 있나?
2. 만일 그렇다면, 그 이유가 연구결과의 p값, 크기 또는 방향 때문인가?

결과의 p값, 크기 또는 방향 때문에 연구결과를 사용할 수 없는 것이 아니라고 가정하는 것이 합리적일 수 있는 시나리오의 예에는 다음과 같은 것들이 있다.

- 관심 중재결과가 연구에서 측정되지 않았음이 명백함(또는 매우 가능성이 높음)(이런 가정은 연구 프로토콜이나 통계분석 계획 확인 혹은 연구자/후원자와의 교신에 근거함).
- 관심 중재결과를 평가하는데 필요한 도구나 기기가 연구 수행 중에 이용 가능하지 않았음.

- 관심 중재결과를 측정했지만 측정기에 문제가 있었거나 상당한 자료 결측으로 인해 분석되지 않았음.

연구결과의 p값, 크기 또는 방향 때문에 연구결과가 결측되었다고 의심하는 것이 합리적일 수 있는 시나리오의 예로는 다음과 같은 것이 있다.

- 중재결과를 측정했다고 주장했지만, 연구결과를 이용할 수 없었고 이에 대한 설명도 없다.
- 관심 중재결과에 대한 그룹 간 차이는 "유의미하지 않은 것"으로 보고된 반면, 다른 중재결과 특히 통계적으로 유의미한 중재군 별 요약 통계치(예: 평균 및 표준편차 같은)는 제시됨.
- 다른 연구결과와 함께 측정되는 경향이 있는 연구결과가 결측됨(예: 특정 원인 사망률에 대해 정보는 있고, 실험 중재군에 더 유리하지만 모든 원인 사망률에 대한 결과가 결측됨).
- 요약 통계치(사건 수 또는 평균 점수)는 전체 그룹에서만 사용할 수 있다(예: 전체 연구대상자 100명 중 10명이 유해효과 사건을 경험했다고 되어 있지만 중재군 별 사건 수는 보고하지 않음).
- 중재결과가 측정된 것으로 예상되며, 연구는 이해상충이 있는 조직에 의해 수행하거나 후원되었다.

중재결과 결측으로 인한 비뚤림위험에 대해 평가될 개별 합성결과에 대한 연구결과의 가용성을 나타내는 매트릭스(연구를 열로, 합성 결과를 행으로)을 구성하는 것이 유용할 수 있다 (Box 8-3)

Box 8-3. fibromyalgia에서 SSRI 효과를 위약과 비교하는 코크란 리뷰 : 세 개의 합성결과에 대한 연구결과의 가용성 메트릭스의 예

			Syntheses assessed for risk of bias		
Study ID	Sample size (SSRI)	Sample size (placebo)	No. with at least 30% pain reduction	Mean fatigue scores (any scale)	Mean depression scores (any scale)
Anderberg 2000	17	18	✓	✓	✓
Arnold 2002	25	26	✓	✓	✓
Goldenberg 1996	22	19	✓	✓	✓
GSK 2005	26	26	✓	–	✓
Norregaard 1995	20	21	✓	✓	✓
Patkar 2007	58	58	✓	–	X
Wolfe 1994	15	9	✓	✓	✓

Key:

✓	A study result is available for inclusion in the synthesis
X	No study result is available for inclusion, (probably) because the P value, magnitude or direction of the results generated were considered unfavourable by the study investigators
–	No study result is available for inclusion, (probably) because the outcome was not assessed, or for a reason unrelated to the P value, magnitude or direction of the results
?	No study result is available for inclusion, and it is unclear if the outcome was assessed in the study

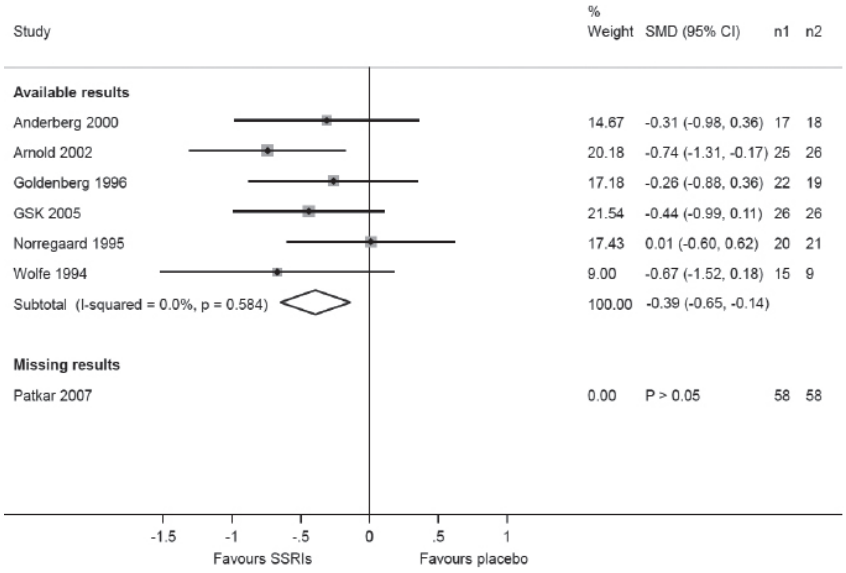
모든 임상시험에 대해 '최소 30% 통증 감소 환자 수' 합성 결과를 얻었다. '평균 피로 점수' 합성 결과에 대해서는, 두 개의 임상시험 연구결과를 이용할 수 없었지만, 결과의 p값, 크기 또는 방향과 무관하였다(이 연구에서 피로는 측정되지 않았다). '평균 우울증 점수' 합성 결과에 대해서는, p값에 근거했을 가능성이 높은 한 연구에 대해 결과를 이용할 수 없었다(임상시험 저자들은 그룹 간에 "의미 있는" 차이가 없다고 보고했으며, 합성에 필요한 자료를 얻으려는 문헌고찰 저자의 시도는 성공하지 못했다). Kirkham 등은 문헌고찰 이득 결과와 위해 결과 둘 다에 대해 템플릿 매트릭스를 적용했다 (Kirkham et al. 2018).

8.3.4 확인된 연구에서 결측된 연구결과 때문에 합성결과가 편향될 가능성이 있는지 고려

결과의 p값, 크기 또는 방향 때문에 일부 연구결과를 사용할 수 없다고 의심하는 경우, 합성결과에 대한 결측의 잠재적 영향을 고려해야 한다. Box 8-4의 예가 도움이 된다.

Box 8-4. 우울증 점수 메타분석에 사용할 수 있는 결과와 누락된 결과를 표시하는 숲 그림

Walitt 등은 Patkar 2007 연구결과의 선택적 비보고를 의심하였다. Walitt 등은 사용 가능한 연구결과와 결측이 있는 연구결과를 분리하여 제시하였다. 사용 가능한 임상시험 표본 크기와 연구결과 결측이 있는 임상시험의 표본 크기를 검사해 보면 Patkar 2007 임상시험의 경우, 모든 적격 임상시험 전체 표본 크기의 거의 1/3이 결측된 수에 해당하였다((58 + 58)/(125 + 119 + 58) = 0.32). 임상시험의 연구결과가 통계적으로 유의미하지 않다는 것을 우리가 알고 있기 때문에, 합성에 포함시키면 요약 추정의 크기가 줄어들 것이다. 따라서 우울증 점수 합성결과에 결측된 연구결과 때문에 비뚤림위험이 있다(Walitt et al, 2015).



선택적 비보고 또는 결과의 과소보고로 인해 합성과정에 누락된 자료의 양이 매우 많은 경우, 그러한 요약 추정치가 심각하게 편향될 수 있다는 것을 근거로 연구의 메타분석을 보고하지 않기로 결정할 수 있다.

8.3.5 추가 연구결과가 합성에서 결측될 가능성이 있는지 평가: "알려지지 않은 무지"

추가 연구결과가 합성에 결측될 가능성을 평가하는 방법에 대해 설명한다.

8.3.5.1 추가 결측 의혹을 제기하는 정량적 신호

연구결과의 결측이 추가될지 여부는 문헌고찰 포함 기준의 정의에 따라 달라진다. 개시 코호트의 연구만 적용 가능하다면, 그 어떤 연구도 누락되지 않을 것이다. MEDLINE의 검색만으로는 모든 관련 연구를 확인할 수 없기 때문에 주제에 따라 CINAHL 및 PsycINFO와 같은 전문 데이터베이스를 검색하지 못하였다면 적합한 연구가 결측될 가능성이 증가할 수 있다. 영어 이외의 언어를 사용하는 국가에서 더 일반적으로 시행되는 중재(예: 전통적인 중의학 중재)을 평가하는 경우, 검색이 영어 논문만 포함된 데이터베이스로 제한될 경우 추가적인 적격 연구가 누락될 가능성이 있다고 가정하는 것이 합리적이다. 이러한 모든 점을 고려하여 이용 가능한 제한된 정보에 근거하여 판단을 내리는 것이 필요하다.

이제 연구결과의 결측 범위에 대해 더 많은 정보를 제공할 수 있는 그래픽 및 통계적 방법에 대해 살펴보기로 하자.

8.3.5.2 깔때기 그림

깔때기 그림은 크기 또는 p값과 관련된 방식으로 메타분석에서 연구결과가 누락될 가능성을 평가하기 위해 오랫동안 사용되어 왔다. 단 세심한 해석이 필요하다. '깔때기 그림'이라는 명칭은 연구 규모가 커질수록 추정된 중재 효과의 정밀도가 높아진다는 사실에서 비롯된다. 따라서 소규모 연구에서 도출된 효과 추정치는 일반적으로 그래프 하단에 더 넓게 산포되며, 더 큰 연구들은 좁게 산포될 질 것이다. 이상적으로는 그림이 대칭으로(뒤집어진) 깔때기와 거의 유사해야 하며 깔때기 그림에 비대칭이 생긴 경우 추가 연구결과가 합성에서 결측될 가능성이 있다(그림 8-1).

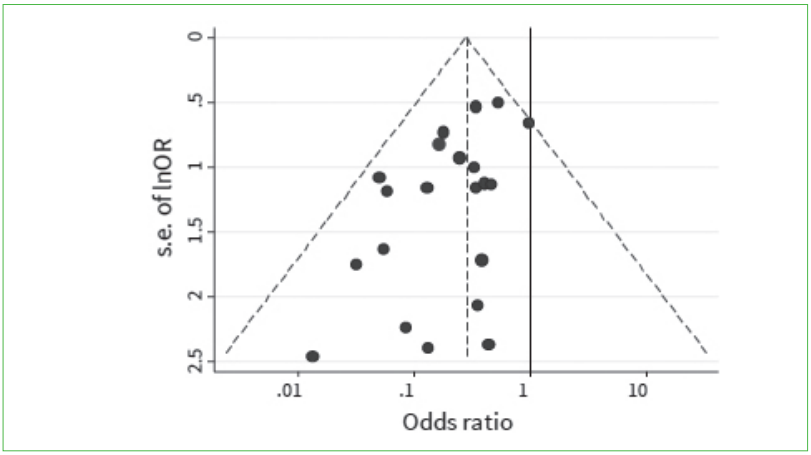


그림 8-1. 연구자료 결측으로 인해 깔때기 그림에 비대칭이 생긴 경우

8.3.5.3 깔때기 그림 비대칭의 다른 이유

깔때기 그림 비대칭성은 오랫동안 비보고 비뚤림과 동일한 것으로 간주되었지만, 깔때기 그림은 소규모 연구 효과의 일반적인 표시 수단으로 보아야 한다. 소규모 연구 효과는 소규모 연구에서 추정된 중재 효과의 경향이 대규모 연구의 그것과 다르다는 것이다. 소규모 연구 효과는 비보고 비뚤림 이외의 다음과 같은 이유로 발생할 수 있다.

- 낮은 방법론적 질 : 방법론적 질이 낮은 연구는 중재효과를 과장하는 경향이 있으므로 실제로 부정적인 결과를 보여야할 연구가 실제로 긍정적인 결과를 보일 가능성이 있으며 이것이 깔때기 그림의 비대칭의 원인이 될 수 있다.
- 이질성 : 고위험군 환자의 경우 이득의 크기가 크고 고위험 환자는 규모가 작은 초기 연구에 포함되는 경우가 많다. 대규모 연구의 중재의 강도가 세지 않아서 효과의 크기가 작은 경향이 있다.
- 특정 효과 추정치의 선택 : 일부 효과 추정치(특히 오즈비와 표준화 평균 차이)는 표준오차와 상관성이 있어서 비대칭을 더 많이 일으킬 수 있다.
- 우연

등고선 강화 깔때기 그림(contour-enhanced funnel plots)은 다른 요인에 의한 비뚤림과 비보고 비뚤림으로 인한 비대칭성을 구별하는 데 도움이 될 수 있다. 등고선 강화 깔때기 그림에는 통계적 유의성의 '이정표(milestones)'에 해당하는 라인이 포함되어 있다. 이를 통해 연구 추정치와 연구결과 결측의 통계적 유의성을 고려할 수 있다. 예를 들어, 연구결과가 통계적으로 유의하지 않고 임상시험 중재에 불리한 영역에서 연구 결측이 의심되는 경우 비대칭이 비보고 비뚤림 때문일 가능성이 높다(그림 8-2).

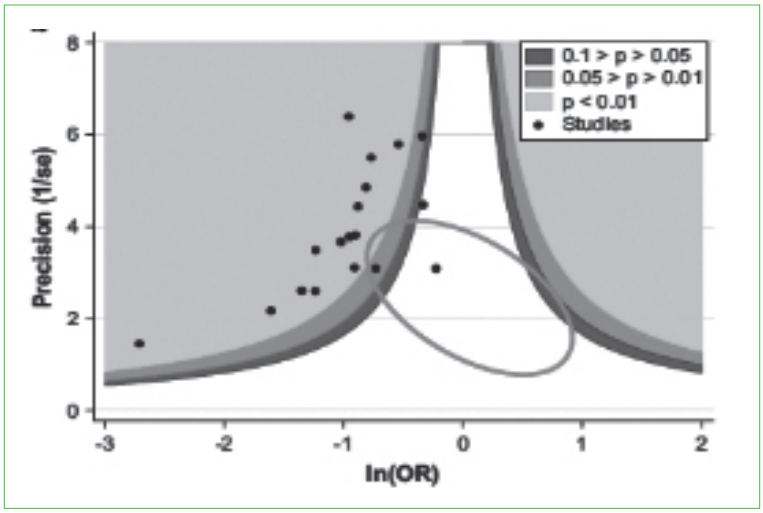


그림 8-2. 등고선 강화 깔때기 그림(contour-enhanced funnel plots).

그림 오른쪽에 누락된 연구 가능성이 있고 이곳은 임상시험 중재에 불리하고 유의하지 않은 영역(즉, P 〉 0.1인 흰색 영역)이기 때문에 비보고 비뚤림일 가능성이 있다.

8.3.5.4 깔때기 그림 비대칭 테스트

깔때기 그림 비대칭성(소규모 연구 효과)에 대한 검정은 추정된 중재 효과와 연구 크기 측정 사이의 연관성이 우연히 발생할 것으로 예상되는 것보다 큰지 여부를 검정하는 것이다. 가장 잘 알려진 검정법은 Egger test이다. 이 검정법은 일반적으로 검정력이 낮으며, 따라서 비보고 비뚤림을 일반적으로 배제할 수 없다. 깔때기 그림 비대칭 검사에는 다음과 같은 권고사항이 있다.

- 깔때기 그림 비대칭에 대한 검정은 10개 이상의 연구가 포함된 경우에만 사용해야 한다.
- 연구들의 규모가 유사한 경우 이 검사를 사용하지 않아야 한다.
- 깔때기 그림 비대칭성에 대한 검정 결과는 깔때기 그림의 육안 검사를 같이 고려하여 해석해야 한다.
- 검정결과 깔때기 그림 비대칭의 증거가 있을 경우, 비보고 비뚤림을 포함한 여러 가지 가능성을 고려하여야 하고 가능한 이유를 구별하려는 시도를 해야 한다.

8.4 비보고 비뚤림 관련 MECIR

C52: 연구결과 누락으로 인한 비뚤림위험 평가(권고)	
체계적문헌고찰이나 체계적문헌고찰에 포함된 메타분석의 결과에서 비보고 비뚤림의 잠재적 영향이 있는지 고려한다.	다양한 유형의 비보고 비뚤림이 있다는 매우 많은 근거가 있다. 비보고 비뚤림은 문헌고찰의 다양한 지점에서 다루어질 수 있다. 철저한 검색과 미발표 연구결과를 얻으려는 노력은 위험을 최소화할 수 있다. 예를 들어 깔때기 그림을 사용하는 등 포함된 연구의 결과를 분석하면 때때로 문제의 가능한 범위를 결정하는 데 도움이 될 수 있다.

8.5 참고문헌과 읽을거리

- Page MJ, Higgins JPT, Sterne JAC. Chapter 13: Assessing risk of bias due to missing results in a synthesis. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA(editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.0(updated July 2019). Cochrane, 2019. Available from www.training.cochrane.org/handbook.
- Eyding D, Lelgemann M, Grouven U, Härter M, Kromp M, Kaiser T, Kerekes MF, Gerken

M, Wieseler B. Reboxetine for acute treatment of major depression: systematic review and meta-analysis of published and unpublished placebo and selective serotonin reuptake inhibitor controlled trials. BMJ. 2010 Oct 12;341:c4737.

- Kirkham JJ, Altman DG, Chan AW, Gamble C, Dwan KM, Williamson PR. Outcome reporting bias in trials: a methodological approach for assessment and adjustment in systematic reviews. BMJ 2018; 362: k3802.
- Walitt B, Urrutia G, Nishishinya MB, Cantrell SE, Hauser W. Selective serotonin reuptake inhibitors for fibromyalgia syndrome. Cochrane Database of Systematic Reviews 2015; 6: CD011735.

제9장 효과 크기 선택

핵심 포인트

- 체계적 문헌고찰에서 흔히 접하게 되는 중재결과 자료의 유형은 이분형 자료, 연속형 자료, 순서형 자료, 계수 자료 또는 비율 자료 및 사건발생 시간(time to event) 자료이다.
- 각 자료 유형에 대한 두 중재군 간의 중재결과 자료('효과 크기')를 비교하는 여러 가지 방법이 있다. 예를 들어, 이분형 중재결과는 위험비, 오즈비, 위험차 또는 필요 치료 수 등을 사용할 수 있다. 연속형 결과는 평균 차이 또는 표준화 된 평균 차이를 사용하여 중재군 비교를 할 수 있다.
- 효과 크기는 상대효과 측정치(예 : 위험비, 오즈비) 또는 절대효과 측정치(예 : 평균 차이, 위험차)이다. 비율 측정은 대개 로그 척도로 분석된다.
- 연구 보고서에서 추출한 연구결과는 분석을 위해 일관성 있고 유용한 형식으로 변환할 필요가 있다.

9.1 자료 유형 및 효과 크기

9.1.1 자료 유형

중재 효과는 일반적으로 아래의 다섯 가지 유형으로 측정 된다.

표 9-1. 메타분석의 주요 자료 유형

자료의 종류	설명
이분형	중재결과가 가능한 두 가지 범주형 응답으로 나타나는 자료
연속형	중재결과가 연속적인 측정값으로 나타나는 자료
순서형	중재결과가 순서를 가지는 범주형으로 표현 되는 자료
계수와 비율	경험한 사건의 수를 나타내는 자료
사건발생 시간 자료	사건이 발생할 때까지의 시간을 나타내는 자료로 중도절단을 포함하는 자료

중재의 영향을 평가할 수 있는 방법은 수집되는 자료의 성격에 달려 있다. 이 장에서는 위의 각 유형의 자료에 대해 중재 효과의 표준 측정치에 대한 정의, 속성 및 해석을 검토하고 학술지 논문과 같은 출처에서 보고될 가능성이 있는 자료로부터 효과 크기를 계산하는 방법에 대

한 팁을 제공한다. 일반적으로 사용되는 효과 크기 및 표준 오차를 추정하는 공식은 Review Manager의 통계 알고리즘 및 기타 표준 텍스트 서적에 나와 있다.

9.1.2 효과 크기(Effect measures)

효과 크기란 두 중재군의 중재결과 자료를 비교하는 통계적 지표이다 예를 들어, 두 그룹 사이의 사건 확률을 비교하는 오즈비와 두 그룹 간의 평균값을 비교하는 평균 차이(mean differences)가 있다. 효과 크기는 비율 및 차이(때로는 상대 및 절대 효과 크기라고도 함)로 크게 나눌 수 있다. 예를 들어, 오즈비는 비율이며 평균 차이는 차이다.

효과 크기(effect estimates)는 결과 자료가 두 그룹 간에 얼마나 다른 지에 대한 중재 효과의 크기를 나타낸다. 비율의 경우 값 1은 그룹 간에 차이가 없음을 나타내고, 차의 경우 값 0은 그룹 간에 차이가 없음을 나타낸다. 이러한 '무효점(null)' 값보다 높거나 낮은 값은 중재를 비교하는 순서(예 : A 대 B 또는 B 대 A)와 결과의 성격에 따라 중재의 이득 또는 위해를 나타낼 수 있다. 모든 추정치는 항상 신뢰 구간(confidence interval)이나 표준 오차(standard error, SE)와 같은 불확실의 측정치와 함께 표현되어야 한다.

9.2 효과 크기의 직접 추출

무작위배정 임상시험의 체계적 문헌고찰에서 일반적으로 각 중재군의 요약 자료는 검색한 문헌에서 자료의 특성에 따라 요약 값을 추출하여 일관된 방식으로 효과를 평가할 수 있도록 하는 것이 좋다. 그러나 때로는 연구 보고서에서 직접적으로 효과 크기를 계산하는 것이 필요하다. 직접 계산이 필요한 경우는 다음과 같다.

1. 특정 유형의 무작위배정 임상시험: 군집 무작위배정 임상시험 및 교차 임상시험 분석의 경우
2. 무작위배정 임상시험의 특정분석: 층화 무작위배정 또는 최소화(minimization)에 사용된 변수를 보정하기 위해 분석을 수행하거나 기저 상태를 보정하기 위해 공분산 분석을 사용한 경우
3. 특정 유형의 중재결과 : 사건발생 시간 자료의 경우와 순서형 및 비율 자료의 경우
4. 비 무작위배정 임상시험: 비무작위배정 임상시험에서 보정된 효과를 추정하는 경우
5. 각 그룹에 대한 요약 자료를 사용할 수 없는 경우: 경우에 따라 각 중재군에 대한 요약 자료를 찾을 수는 있지만 추출 할 수 없는 경우

효과 크기는 신뢰 구간 또는 p-값과 함께 제시 될 수 있다. 일반적인 역변환 가중 평균을 사용하여 메타분석을 수행하는 소프트웨어 절차는 주로 각 연구에서 효과 크기와 표준 오차의 형태로 입력 자료를 취하기 때문에 이러한 수치값에서 표준오차를 얻는 것이 필요하다

9.2.1 신뢰구간 및 p-값을 이용한 표준오차 계산

절대 효과 크기(예 : 표준화된 평균 차이, 위험차, 비율 차이)에 대해 95 % 신뢰 구간(CI)을 사용할 수 있는 경우 표준오차는 다음과 같이 계산 될 수 있다.

$$\frac{|\text{상한} - \text{하한}|}{2 \times 1.96}$$

평균 차이를 추출 할 때 연속형 자료인 경우 신뢰 구간이 정규분포를 사용하여 계산되기 때문에 특별할 주의가 필요하다.

정확한 p-값이 중재 효과 크기와 함께 인용되는 경우 표준오차를 유도는 방법은 아래와 같다.

첫 번째 단계는 표준 정규 분포 표에서 보고된 p-값에 해당하는 Z 값을 얻는 것이다. 표준오차는 다음과 같이 계산 될 수 있다.

$$SE = \frac{\text{효과추정치}}{Z_{p/2}}$$

예를 들어, 초록집 자료에서 위험차 0.03(P = 0.008)을 추정치로 제시한다고 가정하였을 때, 0.008의 p-값에 해당하는 Z 값은 Z = 2.652이다. 위험차의 표준오차는 위험차(0.03)를 Z 값(2.652)으로 나누면 0.011이 된다.

9.3 이분형 자료

9.3.1 이분형 중재결과에 대한 효과 크기

연구대상자의 중재결과가 생존, 임상적 개선여부 같이 둘 중 하나를 선택하는 형태일 때 이분형 자료가 된다. 이분형 자료를 이용한 무작위배정 임상시험에서 가장 흔한 효과 크기는 다음과 같다.

1. 위험비(risk ratio, RR, 상대 위험비라고도 함);
2. 오즈비(odds ratio, OR, 오즈비);
3. 위험차(risk difference RD; 절대 위험 감소(absolute risk reduction)라고도 함)
4. 추가적으로 유익하거나 위해한 결과를 중재하는데 필요 치료 수(number needed to treat, NNT)

표 9-2. 임상연구에서 중재군과 대조군의 연구 결과(이분형 자료)

	일어난 사건의 수('성공')	일어나지 않은 사건의 수('실패')	전체 수
중재군(Experimental, E)	SE	FE	NE
대조군(Control, C)	SC	FC	NC

$$RR = \frac{\text{중재군에서의 사건 발생 위험}}{\text{대조군에서의 사건 발생 위험}} = \frac{S_E/N_E}{S_C/N_C}$$

$$OR = \frac{\text{중재군에서의 사건발생에 대한 오즈}}{\text{대조군에서의 사건발생에 대한 오즈}} = \frac{S_E/F_E}{S_C/F_C} = \frac{S_E F_C}{F_E S_C}$$

$$RD = \text{중재군에서의 사건발생 위험} - \text{대조군에서의 사건발생 위험} = \frac{\text{S}_\text{E}}{\text{N}_\text{E}} - \frac{\text{S}_\text{C}}{\text{N}_\text{C}}$$

$$NNT = \frac{1}{\text{두 군간의 위험도 차이의 절대값}} = \frac{1}{|RD|}$$

9.3.1.1 상대 효과 크기: 위험비 및 오즈비

상대 효과 크기는 한 그룹의 예상 결과를 다른 그룹의 예상 결과와 비교하여 나타낸다.

위험비(RR 또는 상대 위험도)은 두 그룹의 사건 위험의 비율며, 오즈비(odds ratio)는 사건 오즈의 비율이다. 각 비율은 값이 1이면 두 중재의 예상되는 효과가 동일함을 나타낸다.

대조군에 사건이 없는 경우에 대해 위험비나 오즈비를 계산할 수 없다. 중재군의 모든 사람이 사건을 경험하는 경우에는 오즈비를 계산할 수 없다. 이러한 상황과 표준오차가 계산 될 수 없는 상황에서는 2 × 2 표의 각 셀에 0.5를 더하는 것이 일반적이다. 두 군 모두에서 사건이 관찰되지 않는 경우(또는 모든 사건이 관찰된 경우)의 연구는 사건의 상대적 확률에 대한 정보가 없으며 메타분석에서 사용 되지 않는다. 일반적인 사건에서 위험과 오즈가 다르므로 위험비와 오즈비도 다르다. 따라서 오즈비를 위험비로 잘못 해석하면 문제가 발생할 수 있다. 사건의 기회를 증가시키는 중재의 경우, 오즈비는 위험비보다 커질 것이고 그러한 잘못된 해석은 중재 효과를 과대 평가한다. 사건의 기회를 줄이는 중재의 경우, 오즈비는 위험비보다 작을 것이므로, 중재의 효과를 과소 평가한다. 이 해석의 오류는 출판된 개별 연구와 체계적 문헌고찰 보고서에서 흔히 볼 수 있다.

9.3.1.2 절대 효과 측정 : 위험차

위험차(risk difference)는 두 그룹에서 관찰된 위험(관심 결과가 있는 개인의 비율)의 차이이다. 사건이 없는 경우에도 모든 연구에 대해 위험차를 계산할 수 있다. 위험차의 임상적 중요성은 모집단의 사건의 기저 위험의 크기에 따라 다를 수 있다. 예를 들어, 0.02(또는 2 %)의 위험차는 기저 위험이 58 %인 모집단에서는 작거나 임상적으로 유의하지 않은 변화일수도 있고, 기저 위험이 1 %인 모집단에서는 비례하여 훨씬 더 크고 잠재적으로 중요한 변화를 나타

낼 수 있다. 위험차가 상대적 척도보다 더 직접적인 관련 정보를 제공하기는 하지만, 위험차를 해석 할 때 사건의 기저위험과 사건의 결과를 알고 있어야한다. 위험차와 같은 절대 측정치는 중재의 잠재적 이익과 위험을 동시에 고려하여 비교 할 때 특히 유용하다.

9.4 연속형 중재결과 자료

9.4.1 연속형 중재결과를 위한 효과 크기

중재결과가 연속적인 수치이거나 순서형 자료로 범주의 개수가 많은 경우에 해당된다. 연속형 자료의 메타분석에는 평균 차이와 표준화된 평균 차이 두 가지 요약 통계가 일반적으로 사용된다. 이것은 각 개인의 자료가 중재 후 측정과 기저상태 측정치 변화 값에서 얻을 수 있다.

표 9-3. 중재군과 대조군의 연구 결과(연속형 자료)

	기저값(Baseline)	최종값(Final)	변화량(Change)
중재군(Experimental, E) (중재군의 표본크기: N_E)	M_E,baseline SD_E,baseline	M_E,final SD_E,final	M_E,change SD_E,change
대조군(Control, C) (대조군의 표본크기:N_C)	M_C,baseline SD_C,baseline	M_C,final SD_C,final	M_C,change SD_C,change

(M: 평균, SD: 표준편차)

- 평균차(mean difference, MD): 동일한 측정도구로 측정되었을 때, 두 군 간의 평균값의 차이

MDfinal = ME,final - MC,final(중재군과 대조군 각 군의 최종값 평균 차이)
MDchange = ME,change - MC,change(중재군과 대조군 간 군의 전후값 평균 차이)

$$SE(WMD) = \sqrt{\frac{SD_E^2}{N_E} + \frac{SD_C^2}{N_C}}$$

(SE(standard error): 표준오차, WMD(weighted mean difference): 가중 평균차)

- 표준화 평균차(standardized mean difference, SMD): 동일한(중재)결과를 다양한 측정도구로 측정하였을 때, 단일 단위(scale)로 연구결과를 표준화하기 위한 요약통계량으로써 평균차를 연구대상자들에서 관찰된 표준편차로 나눈 값이다.

$$SMD = \frac{\text{각 군간의 해당 (중재)결과의 평균값 차이}}{\text{전체 참여자의 해당 (중재)결과의 표준편차}}$$

$$SE(SMD) = \sqrt{\left[\frac{N_E + N_C}{N_E N_C} + \frac{SMD^2}{2(N_E + N_C)}\right]}$$

9.4.1.1 평균차(mean difference)

평균차는 무작위배정 임상시험의 두 그룹 간의 평균의 절대 차이를 측정하는 표준 통계량이다. 실험 중재가 비교중재와 비교하여 평균적으로 결과를 변화시키는 양을 추정한다. 모든 연구에서 중재결과 측정이 동일한 척도로 이루어진 경우 메타분석에서 요약 통계로 사용할 수 있다.

9.4.1.2 표준화된 평균차(standardized mean difference)

연구가 모두 동일한 중재 결과를 평가하지만 다른 단위로 측정 될 때 표준화 평균 차(SMD)를 메타분석에서 요약 통계량으로 사용한다. SMD는 연구대상자 간 변동(variability)에 대한 효과의 크기를 나타낸다. SMD 방법은 척도의 방향 차이를 고려하지 않는다는 점에 유의해야 한다. 표준화 전에 포함 된 모든 효과의 방향이 같아지도록 해야 한다.

9.4.2 연속형 결과를 위한 자료 추출

연속형 자료의 메타분석을 수행하려면 다음을 찾아야한다.

- 각 중재군의 중재결과 측정의 평균값
- 각 중재군의 중재결과 측정의 표준 편차
- 각 중재군의 중재결과 측정의 연구대상자 수

발표 된 문헌은 연구의 특성에 따라 다양하게 보고를 하기 때문에 제시된 자료 요약에서 이 수치를 얻는 것이 어렵거나 불가능할 수 있다. 보고되는 요약 자료는 평균 또는 중앙값과 표준편차(SD) 또는 표준오차(SE), 신뢰 구간, 사분위수 범위, 범위 등 다양한 통계량이 보고된다. 또한 자료 분석을 위해 선택한 척도도 다를 수 있다. 특히 많이 발생하는 오류는 표준오차를 표준편차로 오해하는 것이다. 불행히도, 어떤 것이 보고되고 어떤 논리적 추론과 다른 연구와의 비교가 요구 될지 명확하지 않다. 표준편차 및 표준오차는 연구 보고서에서 때때로 혼란스럽고 용어가 일관성 없이 사용된다.

때로는 연구대상자 수, 평균 및 표준편차가 보고되어 있지 않지만 대신하여 MD 또는 SMD와 같은 효과 크기가 보고 되는 경우도 있다. 이러한 자료는 표준오차, 95 % 신뢰 구간 또는 정확한 p-값과 같은 불확실성 척도가 수반되는 경우에만 일반 역 분산 기법을 사용하여 메타분석에 포함될 수 있다.

9.4.2.1 중재 후와 기저 자료로부터의 변화량 추출

일반적으로 체계적 문헌고찰에 선정된 연구에 기저 값과 중재 후 값의 변화가 혼합되어 보고 될 것이다. 일부 연구는 둘 다 보고하고 일부는 변화량 또는 중재 후 값만 보고한다. 중재 후 값과 변화량은 때로는 동일한 분석에 결합 될 수 있으므로 반드시 이것이 문제가 되는 것은 아니다. 필요한 평균과 표준 편차를 사용할 수 있는 경우 기저 값과 중재 후 결과의 변화량에 대한 자료를 추출할 수 있다.

기저 값로부터 변화량에 대한 정보를 추출할 때 또 다른 문제는 미방문과 연구 철회로 인해 연구대상자 수가 따른 기저 및 중재 후 값이 보고 될 수 있다는 것이다. 즉, 기저 값은 연구에 포함 된 모든 연구대상자로부터 측정하고 중재 후의 경우 중재 후 값이 있는 연구대상자만을 측정하여 보고 될 수 있다. 변화량을 계산할 수 있는 기저와 중재 후 측정을 모두 보고하는 연구대상자를 식별하는 것은 어려울 수 있다.

표준편차나 표준오차를 추정할 수 없고 사분위수 범위나 범위를 알 수 있는 경우 이를 이용하여 표준편차를 추정하는 경우도 있다. 이 경우는 중재결과의 모집단의 분포가 정규분포라는 가정을 하게 된다. 정규분포의 가정 하에서 사분위수는 약 표준편차의 1.35배가 되기 때문에 이 사실을 이용하여 표준편차를 추정할 수 있다. 또한 같은 가정 하에서 표준편차는 범위의 $\frac{1}{4}$로 추정할 수 있다. 그러나 일반적으로 중재 결과는 정규분포를 따르지 않기 때문에 사분위 범위나 범위를 이용하여 표준편차를 추정하는 것은 권장 하지 않는다.

9.4.2.2 변이에 대한 정보 없음

표준편차가 결측 된 경우에는 임상시험 보고서에서 표준편차를 계산할 수 없으며 표준편차가 결측 된 연구를 제외하지 않을 경우 해당 연구의 결측값을 대치(imputation)해야 한다.

가장 단순한 대치는 하나 이상의 다른 연구에서 표준편차를 빌려 오는 것이다. 여러 후보 표준편차가 사용 가능한 경우 평균, 최고치, '합리적으로 높은' 값 또는 기타 전략을 사용할 것인지 결정해야 한다. MD의 메타분석의 경우, 더 큰 표준편차를 선택하면 연구가 하향가중 되고(down-weights) 더 넓은 신뢰 구간이 산출된다. 그러나 SMD 메타분석의 경우 큰 표준편차를 선택하면 결과가 효과 부족으로 편향될 수 있다.

모든 결측값 대치 방법은 가정을 포함하기 때문에 가능하면 사용하지 않는 것이 좋다. 메타분석에서 분석에 포함 된 대다수의 연구에서 표준편차가 결측 된 경우 대치해서는 안된다. 결측값을 대치한 경우에는 민감도 분석을 통하여 대치값의 영향을 평가할 수 있다.

9.4.2.3 기저 상태로부터의 변화에 대한 표준 편차의 입력

중재 전후 차에 대한 표준편차가 결측된 경우에는 특별한 경우로 결측값을 대치할 수 있다. 아래 표와 같이 중재 전후의 평균과 표준편차만을 사용 가능할 때가 그 경우에 해당한다.

표 9-4. 차이에 대한 표준편차(SD)가 없는 경우

	기저값	최종값	차이
중재군(표본크기)	평균, SD	평균, SD	평균
대조군(표본크기)	평균, SD	평균, SD	평균

각 집단의 평균 차이는 명시적으로 제시되지 않았더라도 기저 평균으로부터 사후 중재 평균을 뺌으로써 얻을 수 있다. 그러나 이 표에 있는 정보로는 차이의 표준편차를 계산할 수 없다. 차이에 대한 표준편차를 계산할 수 있는 충분한 정보가 없을 때, 다른 연구에서 나온 동일한 결과 측정값에 대한 평균 차이 표준편차를 사용하여 정보를 대치 할 수 있다. 그러나 다른 연구의 표준편차를 사용하는 것이 적절하다는 것은 연구가 동일한 측정 척도를 사용했는지, 같은 정도의 측정 오류가 있었는지, 기저 상태와 중재 후 측정의 시간 간격이 같은 지, 그리고 유사한 집단에 있는 지에 달려있다.

상관계수를 이용하여 중재 전후의 결과의 차에 대해 결측된 표준편차를 계산하거나 대치하기 위해 사용될 수 있다. 상관 계수를 이용하는 방법은 연구대상자에서 기저상태 값과 중재 후 측정치가 얼마나 유사한가를 나타낸다. (1) 자세하게 보고된 연구에서 상관 계수를 계산하는 방법과 (2) 상관계수를 이용하여 표준편차를 대체하는 방법을 설명하였다.

(1) 자세히 보고된 연구에서 얻은 상관 계수 계산.

예를 들어 연구에서 기초 및 중재 후('최종') 측정뿐만 아니라 변화를 위한 평균과 표준 편차를 제시한다고 가정 해 보겠다.

표 9-5. 평균과 표준편차가 보고된 자료($평균 \pm SD$)

	기저값	최종값	차이
중재군(표본크기)	15.2±6.4	16.2±7.1	1.0±4.5
대조군(표본크기)	15.7±7.0	17.2±6.9	1.5±4.2

평균과 표준편차를 아는 경우 실험 전 후 차이에 대한 상관 계수는 아래와 같이 계산할 수 있다.

$$Corr_{차이} = \frac{SD^2_{기저} + SD^2_{최종} - SD^2_{차이}}{2 \times SD_{기저} \times SD_{최종}}$$

이 연구에서 최종 열의 자료 만 사용하여 기저 상태로부터의 변화를 분석 할 수 있다. 이 연구에서 다른 자료를 사용하여 각 중재군에 대해 하나씩 두 개의 상관 계수를 계산할 수 있다.

실험군의 상관 계수 $Corr_E$는 다음과 같이 계산할 수 있다.

$$Corr_E = \frac{6.4^2 + 7.1^2 - 4.5^2}{2 \times 6.4 \times 7.1}$$

대조군의 를 얻기 위해서도 마찬가지이다. 이 예에서는

$$Corr_C = \frac{7.0^2 + 6.9^2 - 4.2^2}{2 \times 7.0 \times 6.9}$$

기저 상태 또는 중재 후 표준편차가 이용 가능하지 않을 때, 중재가 결과 척도의 다양성을 변화시키지 않는다고 가정하는 것이 타당하다면, 다른 것으로 대체될 수 있다. 두 중재군의 상관 계수가 서로 비슷하다고 가정하면 단순 평균은 연구의 모든 개인에 대한 기준 및 최종 측정의 유사성을 합리적인 척도로 취할 수 있다(이 예에서는 평균 0.78 및 0.82는 0.80이다). 메타분석에서 다수의 연구에 대해 상관 계수를 계산하고 일관성을 검사 할 것을 권장한다. 상관 계수가 다르면 표본 크기가 너무 작아 안정적인 산정을 할 수 없거나 중재가 결과 측정의 변동성에 영향을 미치거나 중재 효과가 기준 수준에 따라 달라지므로 평균 사용을 피하는 것이 가장 좋다. 또한 0.5보다 작은 값(상관 계수가 -1과 1 사이 임)이 있으면 기저 상태에서의 변화를 사용할 때 별다른 이점이 없으며 중재 후 측정에 대한 분석이 더 정확할 것이다.

(2) 상관 계수를 사용하여 기저 상태로부터의 표준 편차를 대체

이제 기저 상태로부터의 변화의 표준편차가 결측된 연구를 고려하라. 기저 상태와 중재 후 표준편차가 알려질 때, 우리는 상관 계수에 대해 값 Corr을 사용하여 결측 된 표준편차를 대신 할 수 있다. Corr 값은 메타분석에서 다른 연구(위의(1)의 방법을 사용)에서 계산되거나 다른 곳에서 귀속되거나 추론된 논쟁에 근거하여 가정 될 수 있다. 이 모든 상황에서 Corr의 다른 값을 사용해 민감도 분석을 수행하여 분석의 전체 결과가 입력된 상관 계수의 사용에 견고한지 여부를 결정해야 한다.

임상시험적 중재를 위한 기저 상태에서 변화의 표준 편차를 대체하려면 아래의 식을 이용한다.

$$SD_{\text{차이}} = \sqrt{SD_{\text{기저}}^2 + SD_{\text{최종}}^2 - (2 \times SD_{\text{기저}} \times SD_{\text{최종}})}$$

표준편차 중 하나가 이용 가능하지 않다면, 중재가 결과 척도의 변동성을 변화시키지 않는다고 가정하는 것이 합리적일 때 다른 하나가 대체 될 수 있다.

표 9-6. 차이의 표준편차가 보고 되지 않은 자료(평균 ± SD)

	기저값	최종값	차이
중재군(표본크기)	12.4±4.2	15.2±3.8	2.8
비교군(표본크기)	10.7±4.0	13.8±4.4	3.1

위의(1)에서 계산 된 상관 계수를 0.80으로 사용하여 비교군에서 기저 상태로부터의 차이를 다음과 같이 나타낼 수 있다.

$$SD_{중재,차이} = \sqrt{4.0^2 + 4.4^2 - (2 \times 0.80 \times 4.0 \times 4.4)} = 2.68$$

9.4.2.4 평균값 결측(Misssing means)

평균값을 대신하여 중앙값을 대신 사용할 수 있는 경우 자료 분포가 대칭인 경우 평균과 매우 유사하므로 메타분석에서 직접 사용할 수 있다. 그러나 대칭이 아닌 경우 평균과 중앙값은 서로 매우 다를 수 있다. 그럼에도 불구하고, 중앙값이 종종 평균에 대한 합리적인 대치가 될 수 있다.

제1 사분위수 및 중간값 및 제3 사분위수를 이용하여 결측된 평균치를 대치하는 방법이 있으며, 표본 크기, 최소값과 최대값, 사분위 값 및 중간 값을 사용하여 결측 된 평균에 대한 근사값을 도출할 수도 있다.

9.5 순서형 중재결과 자료

9.5.1 순서형 중재결과를 위한 효과 크기

순서형 자료는 각 연구대상자가 한 범주에서 분류 될 때와 그 범주가 자연 순서를 가질 때 발생한다. 범주의 수가 증가함에 따라 순서형 자료는 연속적인 자료와 유사한 특성을 가진다.

측정도구(instrument)는 행동 및 우울증, 인지 능력과 같이 정량화하기 어려운 상태를 측정하는 데 자주 사용된다. 측정도구는 일반적으로 일련의 질문이나 과제를 포함하며 각 과제는 점수가 매겨지며 점수가 합산되어 총 '점수'가 산출된다. 항목이 동등한 중요성을 가지지 않는다면 가중 합계를 사용할 수 있다.

순서 자료와 관련된 방법은 범주 수가 많으면 다루기 힘들어진다. 이런 경우에는 적절한 임계값을 기준으로 범주를 결합하여 이분형 자료로 바꾸어 분석 하거나 앞서 기술한 바와 같이 범주의 수가 많아지면 연속적인 자료와 유사한 성질을 가지므로 연속 자료분석 방법을 이용하여 분석 할 수 있다. 전자는 범주를 결합하기 위한 임계값을 부적절하게 설정 하는 경우 수 비

뚤림이 있으므로 적절한 임계값을 설정하는 것이 매우 중요하다. 순서형 자료를 적절한 임계값을 기준으로 두 그룹으로 분류하여 두 그룹 중 하나는 사건으로 정의하여 효과의 크기는 위험비 또는 오즈비, 위험차를 사용하여 나타낼 수 있다. 순서 자료를 연속적인 자료로 고려하여 분석할 때는 평균 점수는 각 그룹에서 계산되고 효과의 크기는 MD 또는 SMD 추정할 수 있다.

9.5.2 순서형 중재결과에 대한 자료 추출

순서 결과에 대해 추출 될 자료는 순서자료가 분석을 위해 이분법, 연속 결과로 처리 될지 또는 순서 자료로 직접 분석 될지에 따라 달라진다. 자료 추출이 정의된 임계값 상위/하위 연구대상자 수 또는 평균값 및 표준 편차 계산을 포함할지 여부를 사전에 지정하는 것은 불가능할 수 있다. 따라서 자료 추출에 선정된 연구에서 모든 가능한 형태를 자료를 추출하는 것이 좋다. 어떤 경우에는 한 가지 이상의 형태의 분석이 문헌고찰에 포함될 수 있다.

순서 자료를 이분형으로 나누는 경우 임계값 선택에 대한 민감도 분석을 체계적 문헌고찰 연구를 시작하는 단계에서부터 계획하는 것이 좋다. 각 대안적 이분법에 사용될 자료를 수집하려면 문헌에서 자료를 두 번 이상 추출하지 않아도 되도록 짧은 범주를 이용하고 범주에 숫자를 기록해야 한다. 선정된 연구가 조금 다른 순서 척도를 사용하고 공통적인 임계값의 설정 가능성이 명확하지 않은 경우 가능한 모든 임계값을 이용하는 방법로 가능하다.

9.6 계수자료와 비율 자료

9.6.1 계수와 비율 자료의 효과 크기

단순히 각 개인이 사건을 경험했는지 아닌지가 아닌 사건이 발생하는 횟수를 처리하는 것이 바람직하거나 필요할 수도 있다. 이러한 유형의 자료를 계수자료라고 한다. 실제적인 목적을 위해, 계수 자료는 흔치 않은 사건의 계수 및 공통 사건의 계수로 편리하게 분할 될 수 있다.

희귀한 사건의 수는 종종 통계에서 '포아송 자료(Poisson data)'라고 한다. 희귀한 사건의 분석은 발생률(rates)에 초점을 맞춘다. 발생률은 발생 횟수와 관련이 있다. 예를 들어, 한 임상연구의 결과로 314 명의 심근 경색(MI)이 314 명의 추적 관찰 기간 동안 경우 발생률은 일인당 연0.057(0.057 per person-year), 혹은 100인년당 5.7(5.7 per 100 person-years)명이다. 메타분석에서 일반적으로 사용되는 요약 통계는 발생비(rate ratio, RR)로, 두 그룹의 사건 비율을 서로 나누어 비교한다.

임상시험 중재군에서 T_E인년간(T_E person-years) E_E사건이 발생했고 비교군에서 T_c인년 동안 E_C사건이 발생했다고 가정한다. RR은 다음과 같다.

$$RR = \frac{E_E / T_E}{E_C / T_C} = \frac{E_E T_C}{E_C T_E}$$

로그 RR의 대략 표준오차는 다음과 같다.

$$SE_{\ln RR} = \sqrt{\frac{1}{E_E} + \frac{1}{E_C}}$$

사건 없는 경우 각 계수에 0.5의 보정이 추가 될 수 있다.

9.6.2 계수와 비율에 대한 자료 추출

계수형 자료의 추출 방법은 체계적 문헌고찰이 수행 되기 이전에 결정하는 것이 바람직하지만 경우에 따라서는 가능한 자료 형식에 따라 방법이 결정되므로 대부분의 연구가 검토될 때까지 결정할 수 없다. 보고 된 형태로 조사 자료를 추출 할 계획을 세워야 한다.

때로는 사건 수 및 인적 사항에 대한 자세한 자료를 사용할 수 없지만 선택된 문헌에 계산된 결과가 있을 수도 있다. 예를 들어, RR 또는 비율차의 추정치가 제시 될 수 있다. 그러한 자료는 표준오차나 95 % 신뢰 구간과 불확실성 측정을 수반하는 경우에만 메타분석을 수행할 수 있다.

계수형 자료는 각 중재군의 총 사건 수와 위험에 처한 총 인-시간(person time)을 추출 할 수 있으면 계수 자료를 비율로 분석 할 수 있다. 연구대상자의 총 수는 비율 자료 분석에는 필요하지 않지만 연구 설명의 일부로 기록되어야 한다.

9.7 사건발생 시간(time to event) 자료

9.7.1 사건발생 시간 중재결과에 대한 효과 측정

사건이 발생하기 전에 경과한 시간에 관심이 있는 경우 사건발생 시간가 발생한다. 사건발생 시간 자료는 사건이 관찰되지 않은 시간과 해당 기간의 끝이 사건에 해당하는지 아니면 관찰의 끝인지 여부를 나타내는 두 가지 값으로 이루어져있다. 일정 기간 사건에 발생하지 않은 연구대상자는 '중도절단'된다. 사건발생 시간 자료는 재발 또는 퇴원과 같은 사망 이외의 사건을 기반으로 할 수 있다.

사건발생 시간 이분형 자료로 분석 할 수 있다. 이를 위해서는 정해진 일정한 기간 동안 모든 환자의 상태를 알아야한다.

사건발생 시간 자료를 분석하는 가장 적절한 방법은 생존 분석 방법을 사용하고 중재 효과를 위험비로 표현하는 것이다. 위험비는 위험 개념과 유사하지만 순간적 위험을 측정하고 지속적으로 바뀔 수 있다는 점에서 미묘한 차이가 있다. 위험비는 대조군이 중재하지 않고 임상시험을 받는 경우 연구대상자가 특정 시간에 사건에 얼마나 많이(또는 덜) 발생할 가능성이 있는지를 나타낸다. 메타분석에서 중재를 비교할 때, 위험 자체가 지속적으로 변할 수 있음에도 불구하고, 위험비가 추적 기간 동안 일정하다는 가정을 하는 경우가 종종 있다. 이것은 비례위험(hazards) 가정으로 알려져 있다.

9.7.2 사건발생 시간 중재결과에 대한 자료 추출

사건발생 시간 자료의 메타분석은 일반적으로 원래 연구자로부터 개별 환자 자료를 얻고, 위험비 및 통계적 불확실성의 추정치를 얻기 위해 자료를 다시 분석 한 다음 메타분석을 수행한다. 가장 적절한 요약 통계가 제시되지 않는 경우가 많으므로 게시된 논문의 요약 정보를 사용하여 메타분석을 수행하는 것은 종종 문제가 된다.

요약 통계를 이용하여 역 분산(inverse variance) 방법을 사용한 메타분석을 하는 방법은 세 가지 방법이 있다.

첫 번째 콕스 비례위험모형(Cox proportional hazards)을 사용하여 자료를 분석했을 때 사용할 수 있다. Cox 모형은 일반적인 역 분산 메타분석을 수행하기에 충분한 로그 위험비와 그 표준오차를 직접적으로 제시한다.

두 번째 로그 순위 분석 중에 계산된 통계를 사용하여 대략적인 위험비를 추정하는 것이다. 로그 위험비는 $SE = 1/\sqrt{V}$ 인 $(O-E)/V$로 추정된다. 여기서 O는 실험 중재에서 관찰된 사건 수이고, E는 실험 중재에서 예상되는 사건의 로그 순위, O-E는 로그 순위 통계치이고, V는 로그 순위 통계의 분산이다.

세 번째는 발표된 카플란-마이어 곡선(Kaplan-Meier curve)에서 개별 연구대상자 자료를 재구성하는 것이다. 즉 제시된 생존곡선을 다시 수치 자료화하여 위험비 및 표준 오차를 추정하는 방법이다.

9.8 참고문헌과 읽을거리

- Higgins JPT, Li T, Deeks JJ(editors). Chapter 6: Choosing effect measures and computing estimates of effect. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA(editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.0(updated July 2019). Cochrane, 2019. Available from www.training.cochrane.org/handbook.

제10장 메타분석의 준비

핵심 포인트

- 합성은 일련의 선정된 연구에서 나온 자료를 통해 전체적인 근거에 대한 결론을 도출하는 것을 목적으로 하는 과정이다. 여기에는 연구 특성의 합성과 연구 결과의 메타분석이 포함된다.
- 합성에 대한 전반적인 틀을 사용하여 비교 기획, 합성 준비, 합성 시행, 연구결과 해석과 기술에 도움을 받을 수 있다.
- 연구 특성 표는 연구 전반에 걸친 PICO 요소의 검사 및 비교에 도움이 되며, 이러한 특성의 합성과 메타분석을 위한 연구의 그룹화를 용이하게 한다.
- 연구에서 추출한 자료를 표로 만들면 특정 메타분석에 기여하는 연구의 수를 평가할 수 있으며 메타분석이 불가능할 경우 어떤 다른 통계 합성 방법을 사용할 수 있는지를 결정하는 데 도움이 된다.

10.1 개요

합성은 선정된 연구에서 나온 자료를 하나의 근거에 대한 결론을 도출하는 것을 목적으로 하는 과정이다. 일반적 방법은 두 개 이상의 개별연구의 효과크기에 대한 결과의 통계적 합성인 메타분석이다.

선정된 연구의 질평가는 항상 메타분석보다 선행해야 한다. 중재연구의 질평가는 그 내용을 항목별로 분류하여 어떤 연구를 단일 합성으로 그룹화할 수 있는 지 결정하는 데 필요하다. 선정된 연구의 PICO 요소들의 합성은 문헌고찰 결과의 해석을 뒷받침하며 그 자체로 문헌고찰의 중요한 산출물이다. 합성에는 선정된 연구들의 중재와 비교 중재의 특성, 중재가 시행된 모집단과 세팅, 평가된 중재결과 및 전체 근거의 장단점을 포함해야 한다.

앞서 기술한바와 같이 문헌고찰의 서로 다른 단계에서 이루어져야 하는 결정을 이해하는 데 도움이 될 수 있는 세 가지 유형의 PICO가 있다.

- **문헌고찰 PICO**(프로토콜 단계에서 계획됨)는 연구의 적격성(포함할 사항과 문헌고찰에서 제외할 사항)에 기초하는 PICO이다.
- **개별 합성을 위한 PICO**(프로토콜 단계에서 계획됨)는 계획된 비교(중재 및 대조군, 예후 및

모집단의 하위군 포함)를 정의하고, 특정 합성이 어떻게 구조화될 것인지를 결정하며, 필요한 질문을 정의한다.

- **선정된 연구의 PICO**(검토 단계에서 결정됨)는 선정된 연구에서 실제로 조사된 것이다.

이 장에서는 개별 합성을 위한 PICO와 선정된 연구의 PICO에 초점을 맞추어 그룹화 할 수 있는 연구를 결정하는 기준으로 삼는다. 또한 메타분석을 수행하기 전에 수행할 예비 단계를 설명한다.

10.2 합성의 일반적인 단계

표 10-1에는 연구결과를 합성하는 데 사용되는 방법과 관계없이 적용할 수 있는 합성을 위한 일반적인 단계를 나타낸다. 합성 계획은 프로토콜 작성 단계에서 시작해야 하며, 문헌고찰 질문 및 중재군 간의 비교를 계획하는 데 관련된 단계를 설정하여야 한다.. 이러한 단계에는 합성을 위해 중재, 모집단, 결과 및 연구설계의 어떤 특성을 함께 분류할 것인지 규정하는 것이 포함되야 한다.

표 10-1. 결과 합성의 단계

제 1기	***프로토콜 기***
단계 1.1.	비교 내용 설정
제 2기	***선정된 연구에 대해 요약하고 합성을 준비하는 기***
단계 2.1.	각 연구의 특성을 '선정된 연구의 특성' 표로 요약하고, 그 내용 및 기타 특성을 항목별로 검토한다
단계 2.2.	연구를 연구 간의 특성을 비교(예:행렬표)하여 각 비교에서 그룹화 할 수 있을 만큼 유사한 연구 결정
단계 2.3.	합성에 사용할 수 있는 자료를 결정한다
단계 2.4.	프로토콜 계획으로부터의 벗어난 사항이 있다면 계획된 비교나 결과에 대한 수정이 필요한지 또는 새로운 비교가 필요한지 결정한다
단계 2.5.	각 비교에 기여하는 연구의 특성을 통합한다
제 3기	***합성 기***
단계 3.1.	통계 합성(적절한 경우)을 수행하거나 중재 효과에 대한 구조화된 보고를 제공한다.
단계 3.2.	효과의 방향, 효과의 크기, 근거의 확실성, 시험한 중재 및 시험 대상 모집단의 고려를 포함하여 결과를 해석하고 기술한다.

이 장은 주로 합성의 단계 중 제 2기를 다룬다. 문헌고찰에 포함될 연구를 결정하고 자료를 추출한 후, 단계 2.1~2.5를 통해 계획 실행을 시작할 수 있다. 이 과정은 각 연구의 특성을 상

세히 검토한 다음(2.1단계) 어떤 연구가 합성을 위해 그룹화될 수 있을 만큼 유사한지 결정하기 위해 여러 연구 간에 특성을 비교하는 것으로 시작한다(2.2단계). 합성에 사용할 수 있는 자료의 형식에 대한 검토는 다음과 같다(2.3단계). 이 세 가지 단계는 계획된 비교나 결과에 대한 변경이 필요한지 또는 새로운 비교가 필요한지에 대한 의사결정을 알려준다(2.4단계). 이 장에서 마지막 단계는 각 비교에 기여하는 연구의 특성을 합성하는 것이다(2.5단계). 이 장은 합성 전 자료 확인을 위한 실질적인 팁으로 마무리된다.

2.1, 2.2 및 2.5단계는 주로 연구 특성에 대한 정성적 정보의 분석과 합성을 포함한다. 이러한 단계를 수행하는 데 사용되는 과정은 문헌고찰에서 거의 설명되지 않지만, 선정된 연구의 PICO 요소의 특성과 유사성에 대한 많은 주관적 결정이 필요 할 수 있다. 이 절에서 설명하는 예는 이 과정을 보다 투명하게 하기 위한 접근법을 설명한다.

10.3 메타분석의 예비 단계

10.3.1 각 연구의 특성 요약(2.1단계)

합성의 출발점은 각 연구의 PICO 특성을 요약하고 프로토콜에서 미리 지정된 그룹(또는 도메인)대로 이러한 PICO 요소를 분류하는 것이다. 그 내용은 '선정된 연구의 특성' 표에 보고되며, 2.2단계에서 어떤 연구가 합성을 위해 그룹화 될 수 있는지를 결정하는 데 사용된다.

일부 문헌고찰에서, 선정된 연구의 PICO 요소를 설명할 때 각 연구에서 사용된 라벨과 용어는 유지된다. 비교를 위해 PICO와 일치하는 일관되고 널리 이해되는 용어가 있는 영역에는 이 정도면 충분할 수 있다. 그러나 대부분의 영역에서 용어는 가변적이어서 비교를 위해 선정된 각 연구의 PICO와 비교하거나 여러 연구 간에 PICO 요소를 비교하는 것이 어렵다. 연구 전반에서 PICO 요소의 설명을 표준화하면 이러한 비교가 용이하다. 이 표준화에는 비교를 위해 PICO를 기술하는 데 사용되는 라벨과 용어의 적용, PICO 요소의 설명을 구조화하는 것이 포함된다.

10.3.2 그룹화 할 수 있을 만큼 유사한 연구 결정(2.2단계)

선정된 연구의 PICO가 비교를 위해 PICO에 지정된 라벨과 설명을 사용하여 코드화되면, PICO 요소를 여러 연구 간에 비교하고 각 비교 내에서 그룹화할 수 있을 만큼 유사한 연구를 결정할 수 있다.

연구 특성을 표로 만드는 것은 연구 전반에서 PICO 요소를 탐색하고 비교하는 데 도움이 될 수 있으며, 특히 범위가 넓은 문헌고찰, 하나 이상의 PICO 요소에 대한 다양성 또는 많은 수의

연구를 포함하는 문헌고찰에서 중요하다. 연구 특성에 대한 자료는 여러 가지 방법으로 정렬할 수 있으며(예: 비교 또는 특정 PICO 요소에 의한), 표에는 하나 이상의 PICO 요소에 대한 정보가 포함될 수 있다. 최선의 접근법을 결정하는 것은 표를 만드는 목적과 검토 단계에 달려 있다. 연구 특성을 면밀하게 검토하려면 상세한 표가 필요할 것이다. 예를 들어 중재 효과에 대한 효과변경자(effect modifier)의 확인이 필요한 경우 사전 지정된 특성 차이를 확인해야 한다. 문헌고찰이 진행됨에 따라, 이 세부 사항은 PICO 특성에 대한 표준화된 설명으로 대체될 수 있다.

10.3.3 합성에 사용할 수 있는 자료 결정(단계 2.3)

각 비교 내에서 그룹화할 수 있을 만큼 유사한 연구가 결정되면, 다음 단계는 합성에 사용할 수 있는 자료를 조사하는 것이다. 측정 도구와 시간 프레임을 표로 표시하면 다양성의 가능성을 평가할 수 있다. 다양성은 두 가지 형태로 발생한다. 첫째, 동일한 시간 범위 내에서 동일한 중재결과 영역을 측정하는 데 사용되는 복수의 측정 기구에서 측정 되거나, 둘째 동일한 시간 범위 내에서 측정한 여러 시간점에서 측정된다. 그런 다음 다양성을 처리하기 위해 방법은 미리 지정되어야 한다. 문헌고찰에서는 결과 영역 내에서 특정 결과를 선택하기 위한 일련의 의사결정 규칙을 미리 지정해야 한다.

10.3.4 계획된 비교 또는 결과에 대한 수정이 필요한지 또는 새로운 비교가 필요한지 판단(2.4단계)

이전 단계를 통해 계획된 비교를 수정할 필요성이 있는지를 확인하는 단계이다.. 중재의 변이가 매우 중요하다면 중재군을 다시 설정하는 것이 필요할 수 있다. 연구의 수가 적은 경우, 자료가 희소 한 경우 또는 양쪽 모두인 경우에는, 중재, 모집단, 중재결과에 대한 그룹을 다시 설정할 필요가 있다. 예상 시나리오에 대해 다양한 계획을 수립하면 사후 결정의 빈도를 줄일 가능성이 있지만, 모든 시나리오에 대해 계획하기는 어렵다. 사후 변경이 있다면 이에 대한 근거를 보고해야 한다.

10.3.5 각 비교에 기여하는 연구의 특성 합성(2.5단계)

마지막 단계는 통합 효과를 해석하는 데 필수적으로 각 비교에 기여하는 연구의 특성을 합성하는 것이다. 전체 연구에서 주요 PICO 특성에 대한 정보를 통합해야 하며, 미리 지정된 효과변경인자의 특성에서 각 연구가 중요한 차이가 있는지를 확인해야 한다. 연구 특성의 합성에는 GRADE 평가도 중요한데 근거가 문헌고찰 질문에 직접 적용되는지에 대한 판단(비직접성)과 이질성에 대한 가능한 설명(비일관성)을 검토하는데 중요하다.

연구특성을 표로 제시하면 독자들이 제시된 정보에서 패턴을 찾고 식별하는 것을 더 쉽고 빠르게 할 수 있다. 각 비교에 기여하는 연구의 특성을 표로 작성하는 것은 연구의 그룹화에 관한 의사결정의 투명성을 개선하는 동시에 통합 효과에 기여하지 않는 연구가 설명되도록 하는 데 도움이 될 수 있다.

10.4 합성 전 자료 확인

합성에 착수하기 전에, 개별 연구에서 나온 연구 결과가 올바르게 수집되었음을 확인하는 것이 중요하다. 연구에서 보고된 효과의 크기와 방향을 문헌고찰에 제시되는 방법과 비교해야 한다. 연구 보고서의 인쇄 오류, 자료 수집 및 조작의 정확성, RevMan에 자료 입력 등 여러 가지 잠재적 문제를 점검할 수 있는 상당히 간단한 방법이다. 예를 들어, 표준화된 평균 차이의 방향은 우발적으로 잘못될 수 있다. 기본적인 점검은 제시된 자료와 원래 연구에서 이용할 수 있는 자료 간에 동일한 정성적 결과(예: 영향의 방향 및 통계적 유의성)를 확인하는 것이다.

숲 그림의 결과는 동일한 효과 측정 및 통계 모형을 사용하는 경우 원래 보고서의 자료(점추정 및 신뢰 구간)와 일치해야 한다. 만일 다르다면 중재효과에 대한 다른 측정방법 사용, 기저상태 변화값 측정, 사후 중재 측정 단독 또는 기저상태에 대해 조정된 사후 중재 측정, 유사한 중재군 그룹화, 일차 연구 보고서의 분석 단위 오류에 대한 조정 중 다른 선택과 같은 타당한 이유가 있을 수 있다.

10.5 합성의 유형

이 단계에서는 중재군 간의 계획된 비교를 구현하는 데 수반되는 단계를 설명하고 있다. 다음 단계(3단계)는 메타분석을 수행한다. 효과 크기의 메타분석과 그 확장은 많은 장점을 가지고 있다. 그러나 메타분석이 불가능한 상황이 있으며, 이용 가능한 자료를 최적으로 활용하기 위해 다른 통계 합성 방법을 고려할 수 있다. 사용 가능한 요약 및 합성 방법은 표 10-2에 설명되어 있다.

표 10-2 요약 및 합성에 사용할 수 있는 방법

	요약	통계적 합성방법					
방법	설명/표	개수 세기(Vote counting)	p-값 조합	효과의 요약 통계(비모수 통계값)	메타분석	네크워크 메타분석	하위군분석/ 메타회귀분석
다루는 질문 내용	표시된 문장이나 표를 설명적으로 요약함	효과에 대한 근거가 있는가?	적어도 하나의 연구에 근거가 있는가	효과의 범위와 분포는 어떠한가?	공통중재효과는 무엇인가? (고정효과모형) 평균효과는 무엇인가? (변량효과모형)	다중 중재 중 가장 효과적인 중재는 무엇인가?	중재의 효과에 영향을 주는 요인은 무엇인가?
그림의 예시	숲그림(중재 추정치를 결합하지 않고 제시)	Harvest plot Effect direction plot	Albatross plot	Box and whisker plot Bubble plot	Forest plot	Forest plot Network diagram Rankogram plots	Forest plot Box and whisker plot Bubble plot

10.6 참고문헌과 읽을거리

- McKenzie JE, Brennan SE, Ryan RE, Thomson HJ, Johnston RV. Chapter 9: Summarizing study characteristics and preparing for synthesis. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA(editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.0(updated July 2019). Cochrane, 2019. Available from www.training.cochrane.org/handbook.

제11장 메타분석 수행

핵심 포인트

- 메타분석은 둘 이상의 개별 연구 결과의 통계적 합성이다.
- 메타분석의 잠재적 장점은 정밀도 향상, 개별 연구에서 제기되지 않은 질문에 대답할 수 있는 능력, 그리고 상충되는 주장으로 인한 논란을 해결할 수 있는 기회 등이다. 그러나 특정 연구설계, 연구 내 비뚤림, 연구 간 변동 및 보고 비뚤림을 주의 깊게 고려하지 않을 경우 심각하게 오도될 가능성이 있다.
- 개별 연구에서 중재결과의 측정으로 산출되는 자료 유형(예: 이분형 변수, 연속형 변수)을 숙지하고, 중재군 비교에 적합한 효과 측정 방법을 선택하는 것이 중요하다.
- 메타분석은 다른 연구와 비교한 효과 추정값의 가중 평균에 따른 변동을 제시하는 경우가 대부분이다.
- 사건이 없는 연구는 위험비 또는 오즈비에 기여하지 않는다. 희귀한 사건의 경우 Peto 방법은 다른 방법보다 비뚤림이 적고 강력하다는 것이 알려져 있다.
- 연구 간 변동(이질성)을 고려해야 한다. 변량효과 메타분석은 기저 효과가 정규 분포를 따른다고 가정하여 이질성을 허용하지만, 신중하게 해석해야 한다. 변량효과 메타분석으로부터의 예측 구간은 연구 간 변동 범위를 표시하는 데 유용한 방법이다.
- 메타분석을 준비하는 과정에서 많은 판단이 필요하다. 민감도 분석을 사용하여 전체적인 결과가 잠재적으로 영향력 있는 의사결정에 대해 확고한지 여부를 검토해야 한다.

11.1 메타분석 소개

체계적인 문헌고찰에서 중요한 단계는 모든 연구 또는 일부 연구의 수치적 결과를 통합하는 것이 적절한지에 고려이며, 메타분석은 대조 중재와 비교하여 실험 중재의 효과를 합성하여 전체 통계량을 산출한다. 메타분석의 잠재적 장점은 다음과 같다.

1. 검정력과 정밀도를 향상: 상당수 개별 연구는 표본크기가 작기 때문에 작은 효과 크기가 있는 경우 이를 발견해내기 어렵지만 여러 연구가 통합되면 효과를 발견해낼 기회가 더 높아지므로 메타분석을 수행한다.
2. 개별적인 연구를 통해 제기되지 않은 질문의 해결: 서로 다른 특성을 가진 개별 연구들을 포함하여 메타분석을 수행한 경우 개별 연구 사이에서의 중재 효과의 일관성을 조사해볼 수 있

으며, 조사된 효과추정치에 차이가 있는 경우 이에 대한 이유를 설명할 수 있다.

3. 상반된 연구들 사이에서 발생하는 논쟁을 조정하거나 새로운 가설 생성: 결과에 대한 통계적 분석을 통해 상반된 정도를 정식으로 평가할 수 있으며 개별 연구에서의 서로 다른 결과에 대한 이유가 조사되고 정량화될 수 있으므로 메타분석을 수행한다.

물론 통계 합성 방법을 사용한다고 해서 일차 연구보다 검토 결과가 유효하다는 보장은 없다. 더욱이 다른 도구와 마찬가지로 통계적 방법도 잘못 사용될 수 있다.

11.1.1 메타분석의 원리

메타분석에서 일반적으로 사용되는 방법은 다음과 같은 기본 원칙을 따른다.

첫째 개별연구의 요약추정치(summary estimate)를 계산하며, 각각의 연구에 가중치를 준다.

보다 많은 정보를 제공하는 개별 연구에 더 많은 가중치를 부여한다. 일반적으로 가중치로 표본크기를 사용하며, 사건 발생률, 분산의 역수, 연구의 질 등이 사용되기도 한다.

둘째 가중치가 부여된 평균(weighted average)을 계산하여 통합 요약통계량을 산출한다.

통합된 치료효과 크기(pooled treatment effect estimate)

$$= \frac{\sum \text{of (estimate} \times \text{weight)}}{\sum \text{of weights}} = \frac{\Sigma Y_i W_i}{\Sigma W_i}$$

(Y_i : i 번째 연구의 중재효과, W_i : i번째 연구의 가중치)

11.2 메타분석에 대한 일반적인 역분산 접근법

메타분석 절차의 매우 보편적이고 단순한 버전을 일반적으로 역분산법이라고 한다. 이 접근법은 RevMan에서 가장 기본적인 형태로 구현되며, 이분법 및 연속 자료의 많은 메타분석에서 사용된다.

역분산법은 각 연구에 주어진 가중치가 효과 크기의 분산의 역수로 선택되기 때문에 붙여진 명칭이다. 따라서 표준 오차가 작은 대규모 연구에 표준 오차가 큰 소규모 연구보다 가중치가 더 많이 주어진다. 이러한 가중치의 선택은 통합 효과 크기의 불확실성을 최소화한다.

11.2.1 메타분석을 위한 고정효과법

역분산법을 사용한 고정효과 메타분석은 가중평균을 다음과 같이 계산한다.

$$\text{역분산 가중 평균} = \frac{\sum_i Y_i(1/SE_i^2)}{\sum_i (1/SE_i^2)}$$

(Y_i : i 번째 연구의 중재효과, SE_i : 효과추정치의 표준오차)

분석에 필요한 기본 자료는 각 연구에서 도출한 중재 효과와 표준오차의 추정이다. 고정 효과 메타분석은 모든 효과 크기가 동일한 중재 효과를 추정한다는 가정 하에 유효하며, 이는 다양한 의미로 '고정 효과' 가정, '공통 효과' 가정 또는 '평등 효과' 가정으로 언급된다.

고정효과 모형은 각 연구들에 있어 중재효과의 참값은 단 하나만 존재(common true treatment effect)한다는 전제 하에 관찰된 치료효과 값들의 차이는 표본추출의 오차(sampling error) 때문이라는 가정에서 출발하는 모형이다.

$$T_i = \mu + \varepsilon_i$$

(T_i : 관찰된 효과 크기, μ : 중재효과의 참값, ε_i : 표본추출 오차)

다음과 같은 경우에 고정효과 모형을 고려할 수 있다.

1. 메타분석에 통합된 연구들의 인구집단, 중재법 등이 동질하다고 판단되는 경우
2. 메타분석의 목적이 하나의 진실된 효과 크기를 추정하기 위한 경우: 메타분석을 통해 추정된 인구집단의 효과는 다른 동일 인구집단에 일반화시킬 수 있다.(예. 동일한 약제의 동일용량, 동일 투여경로, 엄격히 제한된 환자 대상으로 수행된 위약과 비교된 신약에 대한 5개의 연구들)
3. 메타분석에 통합된 연구들의 수가 매우 적은 경우: 연구간 변동(between-study variance(τ 2))을 추정할 수 없기 때문에 이 경우는 고정효과 모형만이 사용가능하다.

고정효과 모형에서의 가중치는 연구 내 변이만이 포함되며 어떤 추정방법을 사용하느냐에 따라 연구 내 변이를 구하는 방법이 달라진다. 고정효과 모형에서 가중치를 추정할 때 사용하는 방법은 역분산(generic inverse-variance) 추정법, 멘텔-헨젤(Mantel-Haenszel) 추정법, Peto의 추정법이 있다. 대한 설명은 다음과 같다.

$$\text{고정효과 모형에서의 가중치}(weight) = \frac{1}{variance_{\text{연구 내}}}$$

11.2.2 메타분석을 위한 변량효과 방법

역분산 방법의 변형된 형태는 서로 다른 연구들은 관련되어 있지만 서로 다른 중재 효과를 추정한다고 가정하는 것이고, 이렇게 하면 변량효과 메타분석이 생성된다. 가장 간단한 버전은 데르시모니아어와 레어드 방법으로 알려져 있다.

변량효과 모형은 각 연구들에 있어 중재효과의 참값은 단 하나만 존재하는 것이 아니라 각 연구들에 있어 중재법의 효과는 어떤 중재효과 평균 참값(true average treatment effect)을 중심으로 정규분포를 따른다고 가정하는 모형이다. 즉, 각 연구들은 중재효과 평균 참값을 중심으로 퍼져있는 모집단 내 연구들로부터 임의로 추출된 연구들이고, 이 추출된 연구들의 평균값은 평균 참값을 중심으로 정규분포를 따른다는 의미이다.

따라서 각 연구의 중재효과들 간에 관찰되는 변동(variation)은 표본추출의 오차(sampling error, within-study variation, random variation 등으로도 표현됨)와 각 연구들 간의 변동(between-study variation)이 함께 표현된 것이라고 가정한다. 그러나 큰 규모의 연구과 작은 규모의 연구의 결과가 구조적으로 다를 경우 정규성의 가정이 충족되지 못하므로 변량효과 모형은 적절한 효과추정치를 제공하지 못한다.

연구 간의 이질성이 심할 때, 변량효과 모형은 고정효과 모형에 비해 치료효과의 유의성에 대해 좀 더 보수적인 추정치(더 넓은 신뢰구간)를 제공한다. 변량효과 모형은 고정효과 모형에 비해 작은 규모의 연구에 더 높은 가중치를 주고 더 큰 규모의 연구에는 더 낮은 가중치를 주기 때문에 치료효과의 추정치는 두 모형 간에 다를 수 있다. 소규모 연구들의 결과가 출판 비뚤림이나 소규모 연구에서의 연구 내 비뚤림으로 인해 대규모 연구들의 결과와 구조적으로 다를 경우, 변량효과 모형은 이러한 비뚤림을 더욱 악화시킬 수 있다. 따라서 변량효과 모형이 연구간 변동을 고려한다 하더라도 이질성 문제를 해결하기 위한 이상적인 방법은 아니다. 연구간 변동의 추정치가 0인 경우(연구 간 이질성이 없는 경우) 변량효과 모형은 고정효과 모형과 동일한 추정치를 제공한다.

$$T_i = \theta_i + \varepsilon_i = \mu + \zeta_i + \varepsilon_i$$

(T_i : 중재효과 참값, θ_i과 분산 σ^2을 가진 분포로부터 표본 추출된 관찰된 효과 크기, 여기서 중재효과의 참값 θ_i은 평균 μ와 분산 τ^2을 가진 분포로부터 표본추출된 것임)

각 연구들은 동일한 중재효과를 추정하기 보다는 다른 중재효과를 추정하며, 다음과 같은 경우에 변량효과 모형을 고려할 수 있다.

1. 메타분석에 선정된 연구의 인구집단, 중재법 등이 동질하지 않다고 판단될 때
2. 메타분석의 목적이 일반적으로 폭넓은 인구집단에서 일반화시키는 것일 때(예. 연구대상이

나 중재법의 다양성으로 하나의 치료효과 크기를 가정하기 어려운 경우)

변량효과 모형에서의 가중치는 연구간 변이(between studies variance)가 포함된다는 것이 고정효과 모형과 다른 점이며 어떤 추정방법을 사용하느냐에 따라 연구 간 변이를 구하는 방법이 달라진다.

$$\text{임의효과 모형에서의 가중치}(weight) = \frac{1}{variance_{\text{연구 내}} + variance_{\text{연구 간}}(\tau^2)}$$

변량효과 모형은 이질성(heterogeneity)을 해결하지 않는다는 점을 주의해야 한다. 변량효과 모형은 비뚤림이나 교란변수를 보정해주지 않으며 이질성을 일으키는 원인을 해결해주는 것도 아니기 때문에 연구간 이질성이 심할 경우 변량효과 모형으로부터 계산된 효과 크기는 부적절할 수 있다. 그러므로 변량효과 모형이 이질성을 설명하기 위한 것으로 대체되어 사용되어서는 안 된다.

$$weight = \frac{1}{\left[D + \frac{1}{w_i}\right]}$$

분산에 대한 가중치를 부여할 때는 역-분산 추정법과 멘텔-헨젤 추정법 등을 활용하게 된다.

11.2.3 역분산 메타분석 수행

대부분의 메타분석 프로그램은 역분산 메타분석을 수행한다. 일반적으로 사용자는 결과가 이분법적일 때 2×2표와 같은 각 연구의 중재군에서 요약 자료를 제공하거나, 결과가 연속적일 때 각 그룹의 표준 편차 및 표본 크기를 제공한다. 따라서 효과 크기를 계산할 필요가 없으며, 특히 다른 유형의 자료를 대상으로 하는 방법을 사용할 수 있다. 자료를 각 중재군의 요약통계로 편리하게 이용할 수 있을 때, 역분산법을 직접 실시할 수 있다. 예를 들어 추정치와 그 표준 오차는 '일반적인 역분산' 결과 유형으로 RevMan에 직접 입력할 수 있다. 중재 효과의 비율 측정에 대해서는 자료를 자연 로그로 RevMan에 입력해야 한다. 그러나 원래 척도로 결과를 표시하도록 소프트웨어에 지시하는 것은 간단하다. 이를 두 그룹의 표본 크기를 제공하는 열로 보완하거나 대체할 수 있다. 추정치와 표준 오차를 입력할 수 있으면 메타분석에서 높은 수준의 유연성이 생성된다는 점에 유의해야 한다.

표 11-1. 중재군과 대조군의 연구결과(이분형 자료)

	일어난 사건의 수 ('성공')	일어나지 않은 사건의 수 ('실패')	전체 수
중재군(Experimental, E)	S_E	F_E	N_E
대조군(Comparison, C)	S_C	F_C	N_C

역분산 방법을 이용한 가중평균은 다음과 같이 계산된다.

$$\text{역분산 방법에 의한 가중평균} = \frac{\sum \mathrm{Y_i}(1/\mathrm{SE_i^2})}{\sum (1/\mathrm{SE_i^2})}$$

(Y_i: i 번째 연구의 중재효과, SE_i: i 번째 연구의 중재효과의 표준오차)
역-분산 추정법에서의 오즈비와 상대위험도, 위험차에 대한 분산 계산식은 다음과 같다.

$$\text{역분산 추정법에서의 오즈비에 대한 분산, } var(\ln OR) = \left(\frac{1}{S_E} + \frac{1}{F_E}\right) + \left(\frac{1}{S_C} + \frac{1}{F_C}\right)$$

$$\text{역분산 추정법에서의 상대위험도에 대한 분산, } var(\ln RR) = \left(\frac{1}{S_E} - \frac{1}{N_E}\right) + \left(\frac{1}{S_C} - \frac{1}{N_C}\right)$$

$$\text{역분산 추정법에서의 위험차에 대한 분산, } var(\ln RD) = \frac{S_E F_E}{(S_E + F_E)^3} + \frac{S_C F_C}{(S_C + F_C)^3}$$

11.3 이분법 결과의 메타분석

이분법적 결과에 대해 널리 사용되는 메타분석 방법에는 4가지가 있으며, 3가지 고정 효과 방법(Mantel-Haenszel, Peto 과 역 분산)과 변량효과 방법(DerSimonian 와 Laird의 역분산)이 있다. 이 모든 방법은 RevMan에서 분석 옵션으로 사용할 수 있다. Peto 방법은 오즈비만 결합할 수 있고, 다른 세 가지 방법은 오즈비, 위험비 또는 위험 차이를 결합할 수 있다.

한 그룹에서 사건이 없는 경우('zero cell') 일부 방법에서 사용하는 추정 및 표준 오차 계산에 문제를 일으킬 수 있다.

11.3.1 Mantel-Haenszel 방법

자료가 희소할 때, 사건 비율이 낮거나 연구 크기가 작을 때, 역분산 방법에 사용되는 효과 크기의 표준 오차 추정이 어려울 수 있다. 멘텔-헨젤(M-H)방법은 어떤 효과 크기(예: 위험비,

오즈비, 위험 차이)을 사용하느냐에 따라 달라지는 다른 가중치 체계를 사용하는 고정 효과 메타분석 방법이다(Mantel and Haenszel 1959, Greenland and Robins 1985). 이 방법은 사건이 거의 없을 때 더 나은 통계적 특성을 가진다. 이런 상황은 체계적 문헌고찰에서 흔히 볼 수 있으므로, 고정효과 메타분석에서 Ml-H 방법은 일반적으로 역분산법보다 선호된다. 다른 상황에서는 두 가지 방법이 유사한 추정치를 제공한다.

M-H방법에서 오즈비, $(OR_{M-H}) = \frac{weight_i \times OR_i}{weight_i}$,

i번째 연구에서의 오즈비, $(OR_i) = \frac{S_E \times F_C}{F_E \times S_C}$

$weight_i = \frac{1}{var_i}$

여기서,

i 번째 연구에서의 오즈비에 대한 분산 $(var\, OR_i) = \frac{N_E + N_c}{F_E \times S_C}$

i 번째 연구에서의 상대위험도에 대한 분석 분산 $(var\, RR_i) = \frac{N_E + N_C}{(S_E + F_E) \times S_C}$

i 번째 연구에서의 위험차에 대한 분산 $(var\, RD_i) = \frac{N_E + N_C}{N_E \times N}$

95% 신뢰구간 $= e^{\ln(OR_{M-H} \pm 1.96 \times \sqrt{(var\ OR_{M-H})})}$

M-H 방법에서 오즈비에 대한 분산 $(var\, OR_{M-H}) = \left(\frac{F}{2 \times R^2}\right) + \left(\frac{G}{2 \times R \times S} + \frac{H}{2 \times S^2}\right)$

$F = \frac{S_E \times F_C \times (S_E + F_C)}{(N_E + N_C)^2}$, $G = \frac{[S_E \times F_C \times (F_E + S_C)] + [F_E \times S_C \times (S_E + F_C)]}{(N_E + N_C)^2}$,

$H = \frac{F_E \times S_C \times (F_E + S_C)}{(N_E + N_C)^2}$, $R = \frac{S_E + F_C}{N_E + N_C}$, $S = \frac{F_E \times S_C}{N_E + N_C}$

11.3.2 Peto 오즈비 방법

Peto의 방법은 오즈비 조합에만 사용할 수 있다. 역분산 접근법을 사용하지만 로그 오즈비를 추정하는 근사적인 방법을 사용하며, 다른 가중치를 사용한다. Peto 방법은 'O - E' 통계를

합산하는 것이다. 여기서 O는 관찰된 사건 수이고 E는 각 연구의 실험 중재군에서 예상되는 사건 수이다.

로그 오즈비 계산에 사용된 근사치는 중재 효과가 작을 때(오즈비가 1에 가까울 때), 사건이 특별히 흔하지 않고 연구는 실험 및 대조군 그룹에서 유사한 수를 가지고 있을 때 잘 작동한다. 다른 상황에서 그것은 편향된 결과를 나타낼 수 있다. 이러한 기준이 항상 충족되는 것은 아니므로, 메타분석을 위한 기본 접근방식으로 Peto의 방법을 권장하지 않는다.

Peto의 방법을 사용할 때 0 셀 카운트에 대한 보정은 필요하지 않다. 이러한 이유로 이 방법은 사건이 매우 드물 때 잘 수행된다 또한, Peto의 방법은 로그 순위 검정이 사용된 생존자료 분석을 사용한 연구와 이분법 결과 자료와 연구를 결합하는 데 사용할 수 있다.

11.3.3 이분법적 결과에 대한 효과 크기

중재의 영향은 상대적 효과 또는 절대적 효과로 표현될 수 있다. 위험비(상대위험)와 오즈비는 상대적 척도인 반면, 추가 유익성 결과를 다루기 위해 필요한 위험 차이와 수는 절대적 척도다. 사건 발생의 위험비나 사건 발생 없는 위험비을 계산할 수 있다. 이것은 메타분석을 통해 다른 요약 결과를 제공한다.

메타분석에 사용할 요약 통계량의 선택은 세 가지 기준에 따라 달라진다. 첫째 우리는 중재가 적용될 모집단의 메타분석 및 하위군에서 모든 연구에 대해 유사한 값을 제공하는 요약 통계를 원한다. 요약 통계가 일관될수록 중재 효과를 단일 요약값로 표현하는 정당성이 커진다. 둘째 요약 통계량은 유효한 메타분석을 수행하는 데 필요한 수학적 특성을 가지고 있어야 한다. 셋째로 요약 통계는 사용하는 사람들이 쉽게 이해하고 적용할 수 있어야 한다. 이분법 자료에 대한 효과크기 중에서 절대적으로 좋은 방법이 없으므로 경우에 따라 선택해야 한다. 요약 통계량(및 사건 범주 선택)의 선택이 메타분석의 결론에 중요한지 여부를 조사하기 위해 민감도 분석을 수행할 계획을 세우는 것이 좋은 방법이다.

메타분석을 위해 하나의 통계를 사용하고, 보다 쉽게 해석할 수 있는 통계를 사용하여 결과를 다시 표시하는 것이 합리적이다. 메타분석에 오즈비가 사용되는 경우, 위험비로도 다시 표시할 수 있다. 모든 경우에 동일한 공식을 사용하여 신뢰 상한과 하한을 변환할 수 있다.

11.3.4 희귀한 사건의 메타분석

희귀 중재결과의 경우 메타분석만이 의료 중재의 효과에 대한 신뢰할 수 있는 근거를 얻을 수 있는 유일한 방법일 수 있다. 개별 연구는 대개 희귀한 중재결과의 차이에 대한 검정력이 부족하지만, 많은 연구를 사용한 메타분석은 중재가 희귀한 사건의 발생에 영향을 미치는지 여부는 조사하는 충분한 검정력 가질 수 있다. 그러나 메타분석의 많은 방법은 대표본 근사치를 기반으로 하기 때문에 사건이 드물 때는 적합하지 않다. 따라서 메타분석 방법을 선택할 때

주의해야 한다. 사건이 희귀로 분류되는 위험은 특정한 기준을 정할 수 없다. 메타분석에서 비율이 결과인 연구에서 하나 이상의 대조군에서 어떤 사건도 관찰되지 않는 경우도 희귀한 사건으로 볼 수 있다.

11.3.4.1 하나 이상의 대조군에 사건이 없는 연구

개별 연구에서 하나 또는 두 그룹에서 사건이 관찰되지 않을 때 계산 문제가 발생할 수 있다. 역분산 메타분석 방법에는 각 연구에 대한 중재 효과 크기와 표준 오차의 계산이 포함된다. 하나 또는 양군에서 사건이 관찰되지 않은 연구의 경우 0으로 나누는 오류를 산출한다. 대부분의 메타분석 소프트웨어(RevMan 포함)는 자동으로 문제 제로 카운트를 확인하고, 문제가 발생한 2×2 표의 모든 셀에 고정 값(일반적으로 0.5)을 추가한다. 멘텔-헨젤 방법은 포함된 모든 연구에서 동일한 셀이 0인 경우에만 제로 셀 보정을 요구하므로 보정을 자주 사용할 필요가 없다. 그러나 많은 소프트웨어에서 역분산 방법과 동일한 수정 규칙이 멘텔-헨젤 방법에 적용된다. 오즈비 및 위험비 방법은 모든 연구의 모든 군에서 사건이 발생하지 않는 극한 상황에서만 계산 문제를 겪는 Peto 오즈비 방법을 제외하고 차이 방법보다 더 자주 0 셀 보정을 요구된다.

11.3.4.2 희귀 사건에 대한 메타분석 방법의 타당성

시뮬레이션 연구에 의하면 많은 메타분석적 방법들이, 점근적 통계 이론에 의존하고 있어 희귀한 사건들에 대해서는 잘못된 결과가 도출될 수 있다.

사건 위험이 1%를 초과하는 연구에서 가장 좋은 수행 방법은 0셀 보정, 로지스틱 회귀 분석 및 0셀 보정을 하지 않은 멘텔-헨젤 오즈비, 정확 계산방법 등이었다. 이 방법 중 어느 것도 RevMan에서 사용할 수 없다. 희귀 사건의 경우에서 피해야 하는 방법은 역분산법이다.

11.4 연속형 중재결과의 메타분석

연속 자료의 메타분석을 위해 필요한 중요한 가정은 중재결과들이 각 연구에서 각 중재군이 정규 분포를 가지고 있다는 것이다. 이 가정은 매우 큰 연구에서는 문제가 되지 않지만 항상 충족되지 않을 수 있다.

11.4.1 연속형 중재결과에서 효과 크기

연속형 자료를 이용하여 메타분석을 수행할 때는 기본적으로 개별연구들의 중재군과 대조군 각각의 대상 환자 수, 평균, 표준편차에 대한 추출이 선행되어야 한다. 추출된 값을 이용하여 가중 평균차(Weighted Mean Difference, WMD)와 표준화된 평균차(Standardized Mean Difference, SMD)를 계산하며 이를 이용하여 아래의 원리로 메타분석이 수행된다.

가중평균차 이용 시, 아래의 수식을 통해 가중평균차가 계산되어 메타분석에 활용된다.

$$WMD_{summary} = \frac{weight_i \times mean_i}{weight_i}$$

(WMDsummary : 가중평균차의 요약치, weighti : i번째 연구에서의 가중치, meani : i번째 연구에서의 평균)

$mean_i = mean_E - mean_C$ ($mean_E$: 중재군에서의 평균, $mean_C$: 대조군에서의 평균)

$weight_i = \frac{1}{variance_i} = \frac{1}{SD^2}$ (통합추정치의 표준편차 사용)

95% 신뢰구간 $= mean_s \pm (1.96 \times \sqrt{(var_s)})$

여기서, $var_s = \frac{1}{\sum weight_i}$

(variancei : i번째 연구에서의 분산, SD : 통합 추정치의 표준편차, means : 요약 추정치의 평균, vars : 요약 추정치 평균의 분산)

표준화된 평균차 이용 시, 아래의 수식을 통해 표준화된 평균차가 계산되어 메타분석에 활용된다.

$$SMD_{summary} = \frac{weight_i \times SMD_i}{weight_i}$$

(SMDsummary : 표준화된 평균차의 요약치, weighti : i번째 연구에서의 가중치, SMDi : i번째 연구에서의 표준화된 평균차)

$$SMD_i = \frac{\text{각 군 사이에서의 (중재)결과에 대한 평균차(Difference in mean outcome between groups)}}{\text{전체 대상군에서의 (중재)결과에 대한 표준편차(Standard deviation of outcome among participants)}}$$

$weight_i = \frac{1}{variance_i} = \frac{2 \times N_i}{8 + SMD_i^2}$ (통합추정치의 분산사용)

95% 신뢰구간 = $SMD_{summary} \pm (1.96 \times \sqrt{(var_s)})$ 여기서, $var_s = \frac{1}{\sum weight_i}$

(variancei : i번째 연구에서의 분산, SD : 통합 추정치의 표준편차, vars : 표준화된 평균차 요약 추정치의 분산)

(1) 최종값(final value)과 변화량(change value)이 논문마다 통일되지 않고 혼재되어 보고되고 있을 때: 둘 중 한 가지 측정치를 선택하여 메타분석을 수행한다.
(2) 표준화된 평균차의 경우 최종값으로부터 변화량(평균차와 평균차의 표준편차를 재계산하여 변화량으로 통일시켜 메타분석을 수행할 수 있다. 하지만 가능하면 저자로부터 실제 자료를 얻는 방법을 먼저 시도해 보아야 한다.

평균 차이 접근방식의 경우 표준편차를 표본 크기와 함께 사용하여 각 연구에 주어진 가중치를 계산한다. 작은 표준편차를 가진 연구는 상대적으로 높은 가중치를 부여하고 큰 표준편차를 가진 연구는 상대적으로 작은 가중치를 부여한다. 이는 연구 간 표준편차의 변화가 예후 측정의 신뢰성 차이를 반영하는 경우 적절하지만, 표준편차의 차이가 연구 모집단의 결과 변동성의 실제 차이를 반영하는 경우에는 적절하지 않을 수 있다.

표준화된 평균 차이 접근방식의 경우 표준편차는 연구 가중치의 계산뿐만 아니라 단일 척도에 대한 평균 차이를 표준화하기 위해 사용된다. 따라서 표준편차가 작은 연구는 SMD 추정치가 상대적으로 높은 반면 표준편차가 큰 연구는 SMD 추정치를 상대적으로 작게 만든다. 이를 위해 표준편차의 연구 간 변동은 측정 척도의 차이만 반영하고 결과 측정이나 연구 모집단사이의 변동성의 신뢰성의 차이는 반영하지 않는다고 가정해야 한다.

11.4.2 점수 변화의 메타분석

어떤 상황에서는 기저상태에서 변화치에 기초한 분석이 사람 간의 변이를 줄이기 때문에 사후 중재 값의 비교보다 더 효율적이고 효과적이다. 그러나 변화치를 계산하려면 중재결과를 두 번 측정해야하고, 측정 오차가 실제 기저상태 변이 차이보다 클 수 있는 불안정하거나 부정확하게 측정되는 중재결과에 대해서는 효율성이 낮을 수 있다. 기저상태로부터 변화한 값이라는 중재결과는 사후 측정 결과보다 왜곡 분포가 낮은 경우에도 선호될 수 있다. 잘못된 무작위 배정에 대해 '수정'하는 장치로 사용되는 경우도 있지만, 이 방법을 권장하지 않는다.

결과 변수의 기저값을 측정하는 통계적 접근법은 회귀 모형에 공변량 또는 공분산 분석(ANCOVA)이 있다. 이러한 분석은 표준 오류와 함께 중재 효과에 대한 '조정'된 추정치를 산출한다. 이러한 분석은 가장 자주 수행하는 것이 아니지만, 중재 효과에 대한 가장 정확하고 가장 덜 편향된 추정치를 제공하기 때문에 가능한 경우 분석에 포함시켜야 한다. 단, 평균과 표준편차는 각 중재그룹에 대해 별도로 이용할 수 없기 때문에 일반적인 역분산법을 이용한 메타분석에만 포함될 수 있다.

효과 크기가 SMD일 때 사후 중재 값과 변경 점수는 원칙적으로 표준 메타분석 접근법을 사용하여 결합되어서는 안 된다. 표준화에 사용되는 표준편차는 서로 다른 것을 반영하기 때문이다. 사후 중재 값을 표준화 할 때 표준편차는 한 시점의 사람 간 변동성을 반영한다. 표준편차는 변경 점수를 표준화하면 시간 경과에 따른 사람 간 변화를 반영하므로 사람 내 및 사람 간 변동성에 따라 달라지며, 사람 내 변동성은 측정 사이의 시간에 따라 달라질 수 있다.

11.4.3 왜곡된 자료의 메타분석

평균에 기초한 분석은 최소한 대략적으로 정규 분포된 자료와 매우 큰 시험의 자료에 적합하다. 결과의 원래 분포가 비대칭적인 경우, 자료가 왜곡되었다고 한다. 연속적인 결과를 분석할 때 왜곡된 자료의 가능성과 영향을 고려해야 한다. 지우침은 때때로 결과의 평균과 표준편차로부터 진단될 수 있다. 관측된 평균에서 가능한 최저값(또는 가능한 최고값에서 관측된 평균을 뺀 값)을 계산하고 이를 표준편차로 나누어 비율이 2 미만일 경우 지우침이 나타난다. 비율이 1 미만일 경우 치우친 분포의 근거가 강하다. 원본 결과 자료의 변환은 치우침을 상당히 줄일 수 있다.

11.5 이분법적 결과 및 연속적 결과 결합

때때로 동일한 결과에 대한 자료가 일부 연구에서 이분법 자료로, 다른 연구에서 연속 자료로 제시되는 경우가 있다. 예를 들어, 우울증 척도에 대한 점수는 평균 또는 중재 후 어느 시점에 우울해 있던 환자의 백분율로 보고될 수 있다. 이러한 유형의 정보는 이분법화 된 경우 이해하기 쉽고 더 유용하다. 단, 임계치를 결정하는 것이 자의적일 수 있으며, 연속 자료가 이분법 자료로 변환될 때 정보가 손실된다.

이분법 자료와 연속 자료의 조합을 취급하기 위한 세 가지 방법이 있는데 연속형 중재결과에 대해 평균과 표준편차를 입력하는 것, 이분법적 결과에 대해 계수를 입력하는 것, 그리고 모든 자료를 텍스트 형태로 '기타 자료' 결과로서 입력하는 것이 있다.

오즈비를 SMD로 다시 표시하는 통계적 접근법이 있으며, 이를 통해 이분법적 자료와 연속적 자료를 결합할 수 있다. 간단한 접근법은 다음과 같다. 각 중재군에서 기초적인 연속 측정은 로지스틱 분포를 따르고, 결과의 변동성이 실험 및 대조군 연구대상자 모두에서 동일하다는 가정에 따라 오즈비는 다음의 간단한 공식에 따라 SMD로 다시 표시할 수 있다.

$$SMD = \frac{\sqrt{3}}{\pi} ln OR$$

로그 오즈비의 표준오차는 동일한 상수 ($\sqrt{3}/\pi = 0.5513$)를 곱하여 SMD의 표준 오차로 변환할 수 있다. 또한 SMD에 $\pi/\sqrt{3} = 1.814$ 를 곱하여 로그 오즈비로 다시 표시할 수 있다. 메타분석의 모든 연구에 대해 SMD(또는 로그 승산 비율)와 표준 오차를 계산한 후에는 일반적인 역분산 방법을 사용하여 결합할 수 있다. 표준 오차는 해당 자료를 이분법적이고 연속적인 예후 유형 자료로 입력하고 결과로그 오즈비 및 SMD에 대한 신뢰 구간을 표준 오차로 변환하여 모든 연구에 대해 계산할 수 있다.

11.6 서수 결과 및 측정 척도의 메타분석

순서형 자료는 일반적으로 이분형 자료나 연속형 자료로 간주하여 메타분석 한다. 순서형 자료는 1) 순서형 척도의 범주 수가 적고, 2) 각각의 범주별로 발생이 관찰되며, 3) 모든 연구에서 동일한 순서척도가 사용된 경우, 비례오즈 모형(proportional odds model)을 이용하여 메타분석할 수 있다.

비례오즈 모형은 중재효과의 측정치로 비례 오즈비를 이용한다. 예를 들어 3범주의 순서형 자료(1=최상의 향상, 2= 중간정도의 향상, 3= 거의 향상되지 않음)를 이분형 자료(1, 2=성공, 3=실패)로 분류하며, 이때 자료의 두 범주는 동일한 오즈비를 가짐을 가정한다. 그러므로 비례오즈 모형으로부터 계산된 오즈비는 순서에 관계없이 대조군과 비교된 중재군의 성공에 대한 오즈로 해석될 수 있다. 역-분산 방법을 이용하여 로그변환된 오즈비(log OR)와 95% 신뢰구간에서 재계산된 표준오차를 이용하여 메타분석 할 수 있다.

11.7 계수와 비율의 메타분석

비율 자료는 각 연구대상자가 관찰된 시간과 함께 계수를 측정할 경우 발생한다. 이것은 계수되는 사건이 드물 때 특히 적절하다. 계수 자료를 비율로 분석하는 것이 항상 가장 적절한 접근방식은 아니며 실제로 드물다. 그 이유는 다음과 같다.

1. 일정한 기저 위험의 가정은 적절하지 않을 수 있고
2. 통계적 방법은 다른 유형의 자료만큼 잘 개발되지 있지 않다.

계수형 자료나 비율 자료를 메타분석에 사용하기 위해서는 계수형 자료나 비율 자료 즉, 사건이 발생할 위험은 연구대상자에서 시간에 걸쳐 일정하다는 가정이 필요하다. 관찰된 시간에 걸쳐 각각의 연구대상자에서 사건이 발생하며, 이는 율 자료(rate data)가 된다. 이 경우 임상에서 흔하지 않은 가정으로 적절한 접근법이 아닐 수 있다.

$$\text{비율비(rate ratio)} = \frac{E_E/T_E}{E_C/T_C} = \frac{E_E T_C}{E_C T_E}$$

$$\text{SE(rate ratio)} = \sqrt{\frac{1}{E_E} + \frac{1}{E_C}}$$

TE: 중재군의 추적조사기간
TC: 대조군의 추적조사기간

EE: 중재군의 TE 추적기간 동안에 발생된 사건수
EC: 대조군의 TC 추적기간 동안에 발생된 사건수
SE(rate ratio): 율비의 표준오차

사건의 비율비(rate ratio)를 로그 변환시켜 역-분산 방법을 이용하여 메타분석 할 수 있다.

11.8 사건발생 시간 자료(time-to-event)의 메타분석

사건발생 시간 자료의 메타분석에 대한 두 가지 접근법이 있다. 일차 연구에서 추출한 자료 유형 또는 개별 연구대상자 자료의 재분석에서 얻은 자료의 유형에 따라 달라진다.

'O - E' 와 'V' 통계량이 확보되었다면 직접 RevMan에 입력하여 분석할 수 있으며, 이 때는 고정효과 모형을 이용한 메타분석만이 가능하다.

또는, Cox 비례 위험 회귀 모형의 결과에서 로그 위험비과 표준 오차의 추정치를 구한 경우, 일반적인 역분산 방법을 사용하여 연구 결과를 조합할 수 있다.

사건발생 시간 생존시간 자료의 메타분석을 위한 계산식은 아래와 같다.

1. 중재군과 대조군의 HR이나 관측값(O)과 기대값(E)이 보고된 경우

$$HR = \left[\frac{Observed\ events\ research / \log rank\ Expected\ events\ research}{Obseved\ events\ control / \log rank\ Expected\ events\ control} \right]$$

$$V = \frac{1}{(1/Expected\ events\ research) + (1/Expected\ events\ control)}$$

2. 중재군에서 O-E 값과 logrank V가 보고된 경우
(제한점: 연구에서 전체 사건발생 건수가 적을 경우 1의 결과와 다를 수 있음)

$$HR = \exp\left[\frac{\text{관측된 사건 수}(O) - \text{기대되는 사건 수}(E)}{\text{분산}(V)} \right]$$

$$V = \left[\frac{O-E}{\ln(HR)} \right]$$

3. HR과 신뢰구간이 보고된 경우

$$V^* = \left[\frac{\ln(\text{신뢰구간의 상한치}) - \ln(\text{신뢰구간의 하한치})}{2 \times \text{신뢰구간 상한치에 대한 } z\text{점수}} \right]^2$$

(V* = ln(HR)의 분산)
95% 신뢰구간일 경우 z점수=1.96
99% 신뢰구간일 경우 z점수=2.58
90% 신뢰구간일 경우 z점수=1.64

4. HR과 중재군과 대조군의 사건수를 보고한 경우(1:1 무작위배정)

$$V = \frac{\text{중재군에서 관찰된 사건 수} \times \text{대조군에서 관찰된 사건 수}}{\text{관찰된 총 사건 수}}$$

$$O - E = \ln(HR) \times V$$

5. HR과 총 사건수를 보고한 경우(1:1 무작위배정)

$$V = \frac{\text{관찰된 총 사건 수}}{4}$$

$$O - E = \ln(HR) \times V$$

6. HR, 총 사건수, 중재군과 대조군의 무작위배정된 환자 수가 보고된 경우(1:1 무작위배정이 아닌 경우)

두 군에 임의 배정된 이후에 배제된 환자가 있어 as treated 분석이 이루어진 연구라면, HR과 분산은 배정된 환자수가 아닌 분석된 환자수에 근거하여 계산되어야 한다.

$$V = \frac{\text{총 관찰된 사건 수} \times \text{중재군 수} \times \text{대조군 수}}{(\text{중재군 수} + \text{대조군 수})^2}$$

한편, 한 연구에서 ITT분석과 as treated 분석이 모두 제시되었다면, 배정된 환자 수에 근거(ITT분석) 하여 분석하도록 한다.

11.9 이질성

11.9.1 이질성은?

체계적인 문헌고찰에서 연구들 사이의 어떤 종류의 변동성을 이질성이라고 한다.

메타분석에서의 이질성이란 메타분석에서 통합되는 개별연구들의 결과에서의 변동(variation)이 표본추출 오차 이상으로 관찰되어 우연으로 설명할 수 없는 것을 의미한다. 1) 임상적 또는 방법론적 다양성, 2) 우연, 3) 비뚤림이 통계적 이질성을 발생시킬 수 있으며, 일반적으로 통계적 이질성을 이질성이라고 말한다. 이질성은 다음과 같은 이유로 발생할 수 있다.

- 임상적 다양성
- 방법론적 다양성
- 우연(chance): 개별 연구들, 특히 표본크기가 작은 연구나 사건발생률이 낮은 연구들은 우연에 의해 효과 크기가 극단적으로 나타날 수 있다.
- 비뚤림(bias): 연구비 출처와 출판 비뚤림과 같은 개별 연구들의 타당성을 위협하는 비뚤림은 연구들 간의 치료효과 크기를 다양하게 할 수 있다. 예를 들어, 통계적으로 유의하지 않은 결과를 가진 작은 표본크기의 연구들은 통계적으로 유의하거나 효과 크기가 큰 적은 규모의 연구들에 비해 출판되지 않을 위험이 더 큼

이질성은 크게 임상적 다양성(이질성)과 방법론적 다양성(이질성)으로 구분한다.

표 11-2. 이질성의 종류

임상적 다양성(이질성)	방법론적 다양성(이질성)
• 연구된 환자나 인구집단: 질병의 중증도, 진단명, 연령, 성별	• 연구유형과 비뚤림위험의 다양성: 배정순서 은폐, 눈가림과 같은 연구방법에서의 질적 차이
• 사용된 중재법과 임상 세팅: 의료서비스 제공자의 경험이나 교육훈련 정도의 차이	• 치료효과 크기의 방향과 크기의 다양성
• 공통 중재법(co-intervention)	
• 측정된(중재)결과: 연구들 간에(중재)결과를 측정하기 위해 사용된 요약추정치의 차이	
• 추적조사 기간	

이질성이 존재한다면, 일반적으로 통합된 요약 추정치는 해석하기에 곤란하므로 이질성의 원인에 대해 추가적인 문헌고찰이 필요하다.

11.9.2 이질성 확인 및 측정

연구 결과가 서로 일치하는 정도를 고려하는 것이 필수적이다. 이질성의 확인을 위하여 일반적으로 숲그림(Forest plot)이나 L'Abbé plot이 많이 사용된다. 숲 그림은 개별 연구 결과에 대한 신뢰 구간이 중복되지 않는 경우 일반적으로 통계적 이질성의 존재를 나타내고, L'Abbé plot의 경우는 이분형 자료에서 유형한 방법으로 두 군간의 이질성의 원인을 조사하는데 사용된다. 좀 더 공식적으로, 이질성에 대한 통계적 검정을 이용할 수 있다. 이 카이-제곱 검정은 코크란 리뷰의 숲 그림에 포함된다. 그것은 관찰된 결과의 차이가 우연과 호환되는지 여부를 평가한다. 낮은 p-값(또는 그 자유도에 상대적인 큰 카이2 통계량)은 중재 효과의 이질성(실제로 우연을 벗어난 추정치)의 근거이다.

1. 카이제곱 검정법(Q statistics)

이질성 검정을 위한 카이제곱 검정은 각 연구들의 중재효과가 공통중재 효과값으로부터 얼마나 멀리 떨어져 있는지 검정하는 방법이다. 카이제곱 검정은 메타분석의 일반적인 상황에서 연구들이 작은 규모이거나 선정된 연구의 수가 적을 때 낮은 검정력을 가지며, 연구의 수가 많을 때는 중요치 않은 이질성을 발견해낼 수 있기 때문에 해석에 주의를 기울여야 한다. 따라서 p값은 0.05대신 0.10을 통계적 유의수준으로 사용한다.

- 귀무가설: 이질성이 있다는 근거를 찾을 수 없다
- 대립가설: 이질성에 대한 근거가 있다

P- value 〉 0.10 → 귀무가설 채택 → 고정효과 메타분석 수행
P- value 〈 0.10 → 귀무가설 기각 → 변량효과 메타분석 고려

하지만 이 때 p값의 기준을 절대적인 것으로 받아들여서는 안 된다. 변량효과 모형에서 언급한 바와 같이 분석을 계획하는 단계에서 메타분석에 선정된 연구의 인구집단, 중재법 등이 동질하지 않다고 판단될 때, 또는 메타분석의 목적이 일반적으로 폭넓은 인구집단에서 일반화시키는 것일 때 귀무가설이 채택되었을지라도 변량효과 모형을 이용한 메타분석을 고려할 수 있다.

2, Higgin's I^2 statistic

Higgin's I^2통계량은 비일관성을 정량화시킨 통계량이다.

$$I^2 = \frac{Q - df}{Q} \times 100\%$$

(Q : 카이제곱 통계량(Q-statistic), df : Q statistic의 자유도)

해석은 다음과 같이 한다.

① 0% ≤ I^2 ≤ 40%: 이질성이 중요하지 않을 수 있음
② 30% ≤ I^2 ≤ 60%: 중간정도의 이질성이 있을 수 있음*
③ 50% ≤ I^2 ≤ 90%: 실제적으로 이질성이 있을 수 있음*
④ 75% ≤ I^2 ≤ 100%: 상당한 이질성이 있음

* 이질성의 정도의 해석은 다음에 따라 달라진다.

✓ 효과 크기의 방향과 정도: 통계적인 유의성이나 I^2의 결과로 이질성이 인정되어도 효과 크기의 차이가 작고 이득이 크면 이질성이 중요하지 않은 것으로 간주할 수 있다.
✓ 이질성에 대한 근거의 강도(카이제곱 검정법에서의 p값 또는 I^2의 95% 신뢰구간)

11.9.3 이질성 해결 전략

메타분석에 적합한 것으로 간주되는 연구 그룹 중에서 이질성이 확인되는 경우 많은 옵션을 사용할 수 있다.

1. 자료 입력이 맞게 되었는지 확인한다.(특히 연속형 자료에서의 표준편차와 표준오차와 구분 필요)
2. 메타분석의 가정인 동질성을 충족시키지 못하는 경우라고 판단된다면, 메타분석을 수행하지 않는다.(특히 개별연구들의 효과 크기의 방향이 상반적일 경우)
3. 이질성에 대한 원인을 찾기 위하여 사전에 연구계획서에 기술한 연구의 특성에 따른 하위군 분석이나 메타 회귀분석을 수행한다. 하위 분석이 사전에 계획된 것이 아니라면 해석에 주의를 기울여야 한다.
4. 이질성을 무시한 채 고정효과 메타분석을 수행한다. 고정효과 모형의 p값은 모든 연구들에서 치료효과가 없다는 귀무가설에 대한 유의성 검정을 제공한다.
5. 설명되지 않는 연구들 간 이질성을 통합하는 변량효과 메타분석을 수행한다.
6. 이질성은 효과 크기의 부적절한 선택에 의해 발생될 수 있다. 위험차는 대조군의 사건발생률이 비교적 유사한 경우에, 상대위험도와 오즈비는 사건발생률이 비교적 희귀한 경우에 적절하다.
7. 나머지 연구들과 상반되는 결과를 가지는 1-2개의 이상값을 가지는 연구가 이상값의 결과를 나타내는 명백한 이유를 가진다면 메타분석에서 제외될 수 있다.

11.9.4 변량효과 모형에 이질성 통합

변량효과 메타분석 접근방식은 연구들이 서로 다르지만 관련된 중재 효과를 추정한다는 가정을 포함하고 있다. 이 접근방식은 다른 요인에 의해 쉽게 설명될 수 없는 이질성을 다룰 수 있게 해준다. 변량효과 메타분석 모형은 다른 연구에서 추정된 영향이 일부 분포를 따른다는 가정을 포함한다. 이 모형은 실제 또는 명백한 중재 효과가 다른 이유에 대한 지식 부족을 임의적인 것처럼 고려함으로써 나타낸다. 가정된 분포의 중심은 영향의 평균을 기술하는 반면, 그 폭은 이질성의 정도를 기술한다. 일반적인 분포는 정규분포를 가정한다. 특정 분포 가정의 타당성을 확립하는 것은 어려우며, 이것은 변량효과 메타분석의 일반적인 비판이다. 이 분포에 대해 가정된 형태의 중요성은 널리 연구되지 않았다.

변량효과 메타분석은 연구들 간의 이질성을 통합하기 위해 사용될 수 있지만, 이것은 이질성의 원인을 밝히는 철저한 조사 방법은 아니다. 주로 설명될 수 없는 이질성에 대해 일차적으로 시도될 수 있다. 변량효과 메타분석의 연구간 변동의 추정치 tau-squared(τ^2 or Tau^2)는 다음과 같이 계산된다.

$$T^2 = \frac{Q - df}{C} \qquad df = k - 1 \ (df : \text{자유도},\ k : \text{연구 수})$$

$$Q = \sum_{i=1}^{k} W_i Y_i^2 - \frac{(\sum_{i=1}^{k} W_i Y_i)^2}{\sum_{i=1}^{k} W_i}$$

(W_i : i번째 연구의 가중치, Y_i : i번째 연구의 중재효과)

$$C = \sum W_i - \frac{\sum W_i^2}{\sum W_i}$$

이질성이 있는 연구들에서, 변량효과 메타분석은 고정효과 메타분석에 비해 상대적으로 작은 규모의 연구들에 더 많은 가중치를 준다. 변량효과 모형의 분산은 고정효과 모형에서의 표본추출 오차에 추가적으로 연구간 변동을 포함한다. 따라서 변량효과 모형에서의 요약추정치의 신뢰구간은 고정효과 모형에서 보다 대체적으로 넓은 편이다.

- 소규모 연구 영향(small-study effects): 더 작은 규모의 연구들의 결과가 더 큰 연구들의 결과와 체계적으로 다르다면(더 작은 규모의 연구들에서 출판 비뚤림이나 연구 내 변동의 결과로 발생할 수 있음), 변량효과 모형은 비뚤림이 있는 효과추정치일 것이므로 이러한 경우에는 통합된 메타분석을 제시하지 않거나, 작은 규모의 연구를 제외시키는 민감도 분석을 수행해야 한다.

11.10 이질성에 대한 조사

11.10.1 상호작용 및 효과변경

중재 효과는 모집단 또는 중재 특성(용량 또는 지속시간 등)에 따라 달라지는가? 그러한 변이를 통계학자는 상호작용이라고 부르고, 역학자는 효과변경(effect modification)이라고 부른다. 이러한 상호작용을 검색하는 방법에는 하위군 분석 및 메타 회귀 분석이 있다. 모든 방법에는 심각한 단점이 있다.

11.10.2 하위군 분석이란?

연구들 간의 이질성을 해석하기 위해서 동질할 것으로 예상되는 하위군으로 나누어 분석하는 것이다. 하위군 분석은 프로토콜에 사전에 계획되어야 한다. 하지만 분석을 수행할 변수를 사전에 모두 알기는 어렵기 때문에 이질성을 확인한 후 몇몇 변수에 대해서 하위군 분석을 할 수는 있지만 이 경우 해석에 많은 주의를 기울여야 한다. 그러므로 연구수준의 변수(연구설계, 연구의 질, 질환의 중증도 등)를 사전에 구체화한 후에 수행해야 한다. 사전에 구체화된 특성들에서 하위군의 수를 제한하여 통계적인 오류의 가능성을 줄여야 하며, 하위군이 분석되는데 영향을 미치는 연구 결과들을 연구자가 모르게 하여 비뚤림이 있는 결과를 이끌어낼 가능성을 줄여야 한다.

다중 하위군 분석의 결과는 오해의 소지가 있을 수 있다. 하위군 분석은 자연적으로 관찰되며 임의 비교에 기초하지 않는다. 거짓 음성 및 거짓 양성 유의성 시험은 더 많은 하위군 분석을 수행함에 따라 가능성이 빠르게 증가한다. 만약 이러한 결과에 기초하여 결정적인 결론으로 제시된다면, 효과적인 중재가 거부되거나 비효율적인(또는 심지어 해로운) 중재로 처리될 위험이 분명히 있다. 또한 하위군 분석은 향후 연구 방향에 대한 오해의 소지가 있는 권고안을 생성하여 이를 준수할 경우 부족한 자원을 낭비할 수 있다.

'정량적 상호작용'과 '정성적 상호작용'의 개념을 구별하는 것이 유용하다. 정량적 상호작용은 효과의 방향이 뒤바뀌면, 즉 중재가 한 하위군에서는 유익하지만 다른 하위군에서는 해로울 경우에 존재한다. 정성적인 상호 작용은 드물다. 이것은 메타분석의 가장 적절한 결과가 모든 부분군에 걸친 전체적인 효과라는 주장으로 사용될 수 있다. 정량적 상호작용은 효과의 크기가 다양하지만 방향이 다를 때 존재한다. 즉, 중재가 다른 부분군의 다른 정도에 이로운 경우에 존재한다.

11.10.3 하위군 분석의 수행

메타분석은 연구 하위군 내에서뿐만 아니라 하위군 멤버쉽에 관계없이 모든 연구에 걸쳐 RevMan에서 수행할 수 있다. 각 하위군의 메타분석 결과를 고려하여 각 하위군의 효과 크기를 비교하는 것은 유용하다. 이것은 효과의 크기를 비교함으로써 비공식적으로만 이루어져야 한다. 한 하위군에서 이질성에 대한 영향이나 테스트가 통계적으로 유의한 반면 다른 하위군에서는 통계적으로 유의하지 않을 수 있지만 이것이 하위군 인자가 이질성을 설명한다는 것을 나타내는 것은 아니다. 다른 하위군에는 서로 다른 양의 정보가 포함될 가능성이 있고 따라서 효과를 탐지하는 능력이 다르기 때문에 단순히 결과의 통계적 유의성을 비교하는 것은 오해의 소지가 크다고 할 수 있다.

11.10.4 메타 회귀분석

메타 회귀분석은 연구수준의 공변량에 대한 연구 요약추정치의 회귀분석이다. 따라서 분석 단위는 개별 연구이다. 치료효과와 조사된 연구 특성 간에 선형관계가 있다면, 메타 회귀분석은 하위군 분석보다 통계적 검정력이 더 크다.

회귀분석에서 y는 중재 효과 크기인 로그변환된 오즈비(lnOR), 표준화된 평균차(SMD), 가중평균차(WMD), 로그변환된 상대위험도(ln RR), ln HR 등이 되며, 설명변수인 x는 중재효과의 크기에 영향을 미칠 수 있는 연구들의 특성(예. 연구대상자의 일반적인 특성(예. 연령), 대학병원, 의원과 같은 연구가 수행된 임상 세팅(clinical setting), 추적관찰 기간, 배정순서 은폐 여부나 눈가림 수행여부와 같은 비뚤림위험 평가항목, 연구가 수행된 지역적 위치 등)이다.

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_2 + \ldots + \varepsilon$$

(Y_i: i 번째 연구의 효과 크기, β_0 : 추정된 전체 효과 크기, X_i (i=1,2,...k) : 연구의 다른 특성들, ε: 연구간 변동)

해석: 메타 회귀분석에서 구해진 회귀계수는 다른 설명변수에 보정된(예. X_2, X_3) 설명변수(효과를 변화시키는 잠재적 요인; 예. X_1)가 1단위 증가함에 따라 결과변수(중재효과)가 얼마나 변화하는지를 설명한다.

연구들 간의 모든 변동은 회귀모형에 포함되는 공변량으로 설명가능하다고 생각될 때 고정효과 메타 회귀분석이 적절하며, 변량효과 메타 회귀분석은 공변량에 의해 설명되지 않는 연구들 간의 변동인 잔여 이질성(residual heterogeneity)을 허용한다. 그러나 이질성의 원인이 모두 설명가능하다고 가정하는 것은 합리적이지 않기 때문에, 연구수준의 인자에 의해 설명되

지 않는 잔여 이질성(residual heterogeneity)을 허용하기 위해서는 고정효과 메타 회귀분석 대신 변량효과 메타 회귀분석이 수행되어야만 한다.

메타 회귀분석에는 단순 분석과 다중 분석이 있다. 이 중 다중메타 회귀분석에서 설명변수 간의 상관성이 높다면, 공선성의 문제가 발생한다.

다음의 경우 공선성이 있음을 의미하므로 다중 메타 회귀분석은 가능하면 수행하지 않아야 한다.

- 독립변수들 간에 짝지어진 상관성이 큰 경우
- 독립변수를 추가하거나 제외시켰을 때 회귀계수의 변화가 큰 경우
- 회귀계수의 표준오차가 큰 경우

여러 번 분리시켜 수행한 단중 메타 회귀분석이나 한 번에 여러 개의 공변량을 이용한 다중 메타 회귀분석은 위양성 결과를 가져오는 1종 오류의 가능성을 높인다. 이 때 permutation test로 구해진 통계량을 다중 검정(multiple testing)을 보정하기 위해 사용한다.

Permutation test는 다중 메타 회귀분석에서 p값을 계산하기 위한 접근법으로 회귀계수는 0이라는 귀무가설 하에서 자료를 시뮬레이션하는 비모수적 방법이다. 이는 임의 permutation에 근거한 Monte Carlo simulation을 토대로 한 방법이다.

메타 회귀 분석은 Stata 통계 패키지에 사용할 수 있는 'metarg' 매크로를 사용하거나, 다른 패키지와 함께 R에 대한 'metafor' 패키지를 사용할 수 있다.

11.10.5 하위군 분석 및 메타 회귀분석의 해석

하위군 분석 및 메타회귀분석에 대한 적절한 해석은 주의가 필요하다.

1. 하위군 비교는 관찰적이다.

하위 분석과 메타 회귀분석은 본질적으로 전적으로 관찰적이라는 것을 기억해야 한다. 이러한 분석은 연구 간의 차이를 조사한다. 임상시험 내에서 개인이 한 집단이나 다른 집단으로 임의로 분류된다 하더라도, 그들은 한 실험이나 다른 실험에 들어가기 위해 임의로 분류되지 않는다. 따라서 하위군 분석은 다른 연구 수준 특성에 의한 교란 요인을 통한 가능한 비뚤림을 포함하여 관찰연구의 한계를 가진다. 더욱이, 하위군 간의 진정한 차이조차도 반드시 하위군의 분류에 기인하는 것은 아니다. 예를 들어 백혈병 치료를 위한 골수이식의 하위군 분석 결과에서 형제 기증자의 나이와 이식 성공 사이에 강한 연관성을 보여줄 수 있다. 하지만, 이것은 아마도 기증자의 나이가 중요하다는 것을 의미하지는 않을 것이다. 사실, 공여자 나이는 아마

도 핵심적인 요소일 것이고 하위군 분석 결과는 공여자의 나이와 형제자매의 나이 사이의 강한 연관성 때문일 것이다.

2. 분석은 사전 지정되었는가, 사후 분석되었는가?

저자는 하위군 분석이 사전 지정되었는지 아니면 연구 결과가 작성된 후 수행되었는지 기술해야 한다. 사전 지정된 분석인 경우 하위군 분석을 더 신뢰할 수 있다. 이질성을 설명하기 위해 수많은 사후 하위군 분석을 수행하는 것은 자료 준설(data dredging)의 한 형태다. 자료 준설은 대개 많은 다른 특성을 고려함으로써 이질성에 대한 잘못된 설명을 찾을 수 있기 때문에 비난 받는다.

3. 조사 결과를 뒷받침하는 간접적인 근거가 있는가?

하위군 간의 차이는 설득력이 있는 경우 임상적으로 타당해야 하며 다른 외부 또는 간접 근거에 의해 뒷받침되어야 한다.

4. 차이의 크기가 실질적으로 중요한가?

하위군 간 차이의 크기로 인해 다른 하위군에 대해 서로 다른 권고를 하는 것은 아닌 경우, 전체 분석 결과만 제시하는 것이 더 나을 수 있다.

5. 하위군간에 통계적으로 유의미한 차이가 있는가?

서로 다른 상황에서 중재의 효과가 다른지를 규명하기 위해, 서로 다른 하위군에서 영향의 크기를 직접 비교해야 한다. 특히 별도의 하위군 분석 내에서 결과의 통계적 유의성을 비교해서는 안 된다.

6. 분석은 연구내 또는 연구간 관계를 고려하고 있는가?

환자 및 중재적 특성의 경우, 연구 내에서 관찰되는 하위군의 차이는 연구의 하위 집합의 분석보다 더 신뢰할 수 있다. 이러한 연구 내 관계가 여러 연구 간에 재현될 경우 이는 결과에 자신감을 더한다.

11.10.6 기저 위험(underlying risk)의 영향 조사

일련의 연구들 중에서 잠재적으로 중요한 이질성의 중요한 원천중 하나는 중재결과 사건의 근본적인 평균 위험이 연구들 사이에 다를 때이다. 특정 사건의 기저 위험은 연령이나 질병 심각도와 같은 환자 혼합 요인의 종합 척도로 볼 수 있다. 일반적으로 각 연구의 대조군 에서 사건의 관측된 위험(대조군 위험)으로 측정한다. 기저위험은 알려진 위험 요인과 알려지지 않은 위험 요인의 합으로 볼 수 있기 때문에 이 개념은 임상실무 적절성 면에서 논란이 있다. 또한

대조군 위험이 추적관찰 기간의 길이에 따라 달라지기 때문에 문제가 발생하며, 이는 종종 연구마다 다르다.

서로 다른 요약 통계(위험 비율, 오즈 비율 및 위험 차이)를 사용하면 기저위험과의 다른 관계를 입증할 수 있다. 기저위험과 밀접한 관계가 없음을 보여주는 요약 통계는 일반적으로 메타분석에서 사용하는 것이 바람직하다.

효과 크기와 대조군 위험 사이의 관계를 조사하는 것은 평균에 대한 회귀라고 알려진 기술적 현상에 의해 복잡하다. 이는 대조군 위험이 효과 크기의 필수적인 부분을 형성하기 때문에 발생한다. 전적으로 우연한 기회에 관측된 대조군에서 높은 위험은 평균적으로 기대 효과 크기보다 더 높은 결과를 낳으며, 그 반대의 경우도 마찬가지일 것이다. 이 현상은 효과 크기와 대조군 그룹 위험 사이의 잘못된 상관관계를 초래한다. 평균으로의 회귀 분석을 위한 정교한 소프트웨어가 필요한 방법이 있다. 이러한 분석은 이러한 분석에 사용되어야 하며, 통계적 전문지식이 권장된다.

11.10.7 용량 반응 분석

메타 회귀의 원리는 중재 효과와 사용량(일반적으로 dose-response), 치료 강도 또는 치료 지속시간 사이의 관계에 적용될 수 있다. 연구대상자가 연구 내에서 용량이나 다른 용량에 임의로 추출되고 유사한 연구에서 일관된 관계가 발견되는 경우 용량(또는 유사한 요인)의 차이로 인한 효과 차이에 대한 결론들은 더 강력한 근거가 된다. 저자들은 이러한 효과를, 특히 이질성에 대한 가능한 설명으로 고려해야 하지만, 연구 간 차이에 기초한 결론을 도출하는 데는 신중해야 한다. 저자는 많은 메타 회귀 분석의 낮은 검정력을 고려할 때 dose-response 관계가 존재하지 않는다고 주장하는 데 특히 주의해야 한다.

11.11 결측값

11.11.1 결측값의 형태

체계적인 문헌고찰이나 메타분석에는 많은 잠재적 자료 결측의 출처가 있다. 예를 들어, 전체 연구가 문헌고찰에서 결측될 수 있고, 중재결과가 연구에서 결측될 수 있으며, 중재결과에 대한 요약 자료가 결측될 수 있으며, 개별 연구대상자가 요약 자료에서 결측될 수 있다.

표 11-3. 메타분석에서 결측자료의 형태

결측 자료의 형태	자료 결측의 가능한 원인
결측된 연구	출판 비뚤림 검색이 충분히 포괄적이지 않음
결측된 중재결과	측정되지 않은 결과 선택적 보고 비뚤림
요약 자료 결측	선택적 보고 비뚤림 불완전한 보고
개인의 결측	배정된 대로 분석을 하지 못함 탈락 선택적 보고 비뚤림
연구 수준 특성이 결측됨(하위군 분석 또는 메타 회귀 분석의 경우)	측정되지 않은 특성 불완전한 보고

11.11.2 결측값 처리에 대한 일반 원칙

결측을 처리하기 위한 주요 옵션은 다음과 같다.

1. 사용 가능한 자료만 분석(즉, 결측된 자료 무시)
2. 결측된 자료를 대체 값으로 대체하고 이를 관찰한 것처럼 처리(예: 마지막 관찰 결과, 모든 결과가 불량한 것으로 가정하고 평균을 나타냄, 회귀 분석의 예측 값에 기반한 예측 결과).
3. 결측된 자료를 대체시키고 이러한 자료가 대체되었다는 사실을 고려한다. 불확실성(예: 복수 대체, 단순 대체 방법(포인트 2)과 같은 경우 표준 오류에 대한 조정.
4. 통계적 모형을 사용하여 결측된 자료를 허용하고, 그 자료에 대한 사용 가능한 자료와의 관계를 가정 함

옵션 2는 대부분의 상황에서 실용적이며 체계적인 문헌고찰에 매우 흔히 사용된다. 그러나 일반적으로 너무 좁은 신뢰 구간에서는 대체 값과 결과의 불확실성을 인정하지 않는다.

결측된 자료를 처리하기 위한 5가지 일반적인 권고사항은 다음과 같다.

- 가능하면 원연구자에게 연락하여 결측된 자료를 요청하라.
- 결측 자료를 처리하는 데 사용되는 모든 방법의 가정을 명시한다. 예를 들어, 자료가 임의로 결측된 것으로 가정하거나 결측값이 불량한 결과와 같은 특정 값을 갖는 것으로 가정했다.
- 임의 배정된 시험의 결과 자료 결측으로 인한 비뚤림위험을 평가하려면 8장의 지침을 보라.

- 민감도 분석을 수행하여 가정이 합리적으로 변경되는 데 얼마나 민감한 결과를 도출하는지 평가한다.
- 고찰에서 결측된 자료가 검토 결과에 미칠 수 있는 영향을 다룬다.

11.12 메타분석에 대한 베이지안적 접근법

베이지안 통계는 유의성 검사와 신뢰 구간에 대한 다른 철학에 기초한 통계에 대한 접근이다. 그것은 본질적으로 근거의 갱신에 관한 것이다. 베이지안 분석에서 초기 불확실성은 관심 수량에 대한 사전 분포를 통해 표현된다. 현재 자료 및 생성 방법에 관한 가정은 가능성에 요약되어 있다. 그런 다음 관심 수량에 대한 사후 분포는 사전 분포와 가능성을 결합하여 얻을 수 있다. 가능성은 메타분석(예: 무작위배정 임상시험의 2×2표)과 메타분석 모형(예: 고정 효과 또는 변량효과 가정)에 선정된 연구의 자료를 모두 요약한다. 분석 결과는 보통 각 관심 수량에 대한 사후분포에서 95% 신뢰할 수 있는 구간으로 제시되며, 이는 고전적인 추정치와 신뢰 구간과 매우 유사하다. Bayesian 분석은 WinBUGS를 사용하여 수행할 수 있다.

메타분석을 위한 고전적인 방법에 비해 베이지안적 접근법의 몇 가지 잠재적 장점은 다음과 같다.

- 중재 영향 또는 연구 중 변동 가능 범위와 같은 외부 근거를 통합한다.
- 다양한 임상 결과 상태의 개념을 통합하여 메타분석을 의사결정 맥락으로 확장한다.
- 연구 간 분산 추정치의 부정확성을 자연스럽게 허용한다.
- 기저위험과 치료 유익성 사이의 관계를 조사한다.
- 복잡한 분석 수행
- 자료가 사람들의 믿음을 얼마나 변화시킬지 검토한다.

11.13 민감도 분석

민감도 분석은 메타분석 연구결과의 강건함(robustness)을 조사하기 위해 수행되는 분석 방법이다. 추정치가 강건하면 체계적 문헌고찰의 결과에 더욱 확신을 줄 수 있다. 가정이 달라짐에 따라 연구결과가 달라진다면, 메타분석 결과의 신뢰성이 떨어지므로 그 원인을 찾고 다른 가정에 따라 민감도 분석을 수행하는 등의 적절한 조치를 취해야 한다. 그 이유가 연구 특성에 따른 것이라면 연구 특성들(예. intention to treat 분석의 적절성 등)에 따라 어떤 일차 연구들을 포함 혹은 배제시켜 분석하고 결과의 차이를 관찰해볼 수 있다. 서로 분석결과가 다른 경우 높은 수준의 질을 가진 연구의 통합 추정치에 초점을 두어 결론을 내려야 한다. 한편, 이질성의 원인이 1-2편의 예외적인 연구(이상치, outlier)들 때문이라고 판단되면, 민감도 분석의 일부로서 이상치를 포함 혹은 배제시킨 결과를 제시할 수 있다.

민감도 분석 수행 시 체계적 문헌고찰의 결과에 큰 영향을 준다고 판단되는 결측 정보를 확인하였다면, 불확실성을 해결하기 위해 개별연구자를 접촉하거나 개별환자 자료를 확보해야 한다. 불확실성이 해결되지 않았다면 결과는 적절한 정도로 주의 깊게 해석되어야 하며 추후 연구에 대한 제안을 해야 한다.

연구 검색:

1. 후속 간행물에서 결과를 확인할 수 없는 초록을 검토 대상에 포함시켜야 하는가?

자격 기준:

1. 연구대상자의 특성: 연구의 모든 사람이 연령대를 충족하지 못하는 경우, 연구를 포함시켜야 하는가?
2. 중재의 특성: 메타분석에는 어떤 범위의 용량이 포함되어야 하는가?
3. 대조군 특성: 대조군으로 사용할 통상적인 관리를 정의하기 위해 어떤 기준이 필요한가?
4. 중재결과의 특성: 어떤 시점 또는 시점 범위를 포함할 수 있는가?
5. 연구설계: 맹검 및 비맹검 결과 평가를 포함시켜야 하는가, 아니면 방법론적 기준의 다른 측면에 의해 연구 포함이 제한되어야 하는가?

어떤 자료를 분석해야 하는가?

1. 사건발생 시간 자료: 중도절단된 자료의 분포에 대해 어떤 가정을 해야 하는가?
2. 연속 자료 : 표준편차가 결측된 경우, 언제 어떻게 대체되어야 하는가? 분석은 점수 변화값에 기초해야 하는가, 아니면 중재 후 최종값에 기초해야 하는가?
3. 순서 척도: 짧은 서수 척도를 두 그룹으로 이분화하려면 어떤 절단점을 사용해야 하는가?
4. 군집 무작위배정 임상시험: 임상시험 분석에서 군집화에 대해 보정되지 않은 경우, 군내상관계수(intraclass correlation coefficient) 중 무엇을 사용해야 하는가?
5. 교차 임상시험: 1차 보고서에서 사용할 수 없는 경우 대상 내 상관 계수(within-subject correlation coefficient) 중 어떤 값을 사용해야 하는가?
6. 모든 분석: 결측된 결과에 대해 어떤 가정을 해야 하는가? 보정 또는 보정되지 않은 중재 효과 크기를 사용해야 하는가?

분석 방법:

1. 분석에 고정 효과 또는 변량효과 방법을 사용해야 하는가?
2. 이분법 결과의 경우 오즈비, 위험비 또는 위험차이 중 무엇을 사용해야 하는가?
3. 여러 척도가 동일한 차원을 평가한 연속 결과의 경우, 모든 척도에 걸쳐 표준화된 평균 차이 또는 각 척도에 대해 개별적으로 평균 차이로서 결과를 분석해야 하는가?

11.14 메타분석 수행 관련 MECIR

C65: 치우친 자료 다루기(권고)	
연속형 자료를 분석할 때는 치우친 자료의 함의와 가능성을 고려한다.	치료친 자료는 일반적으로 평균과 표준편차로 요약되지 않는다. 대규모 표본의 경우 통계학적으로 적절하지만, 치우친 자료는 연구가 소규모인 경우에는 잘못된 연구 결과를 도추할 수 있다.
C63: 통계적 이질성 평가하기(필수)	
메타분석 수행시에 연구간 변이의 존재여부와 그 정도를 평가한다.	이질성의 존재는 일반화 가능한 결론이 가능한가에 영향을 준다. 충분한 정보가 있는 경우 이질성을 확인하는 것이 중요하다. 연구가 수가 적을 때는 I^2 및 Tau^2와 같은 측정 값에 많은 불확실성이 있을 수 있다. 따라서 이질성을 진단하기 위해 간단한 임계 값을 사용하는 것은 피해야 한다.
C69 : 결과 해석시 통계적 이질성 고려(필수)	
결과를 해석 할 때, 특히 효과 방향에 변이가 있는 경우 통계적 이질성을 고려한다.	이질성의 존재는 일반화 가능한 결론이 가능한가에 영향을 준다. 고정 효과 분석을 사용하는 경우 신뢰구간은 이질성 범위를 무시한다. 변량효과 분석을 사용하는 경우 연구결과는 여러 연구의 평균 효과와 관련이 있다. 두 경우 모두 이질성의 의미를 주의 깊게 다루어야 한다. 충분한 연구가 있는 경우 이질성의 이유를 이해하는 것이 가능할 수 있다.
C67 : 하위군 비교(필수)	
하위군 분석을 수행하려고 하고, 충분한 연구가 있다고 판단되는 경우 공식적인 통계 검정을 사용하여 비교한다.	하위군간에 통계적 유의 수준의 차이가 있다고 실제로 하위군간 효과에 차이가 있다고 결론을 내리는 것은 오류의 위험성이 높다.
C68 : 하위군 분석 해석(필수)	
하위군 분석이 수행되는 경우, 특정 결과에 대해 과도하게 강조하면 안 되며 프로토콜에 지정된 하위군 분석 계획을 따른다.	특정 하위 그룹 혹은 특정 하위 그룹 분석에 대해 선택적으로 보고하거나 과대 해석을 피해야한다. 여러 하위 그룹 분석을 수행 할 때 특히 문제가 된다. 이는 합리적이고 정직한 하위군 분석을 배제한다는 의미는 아니다.
C64 : 중재결과 결측 자료 처리(권고)	
개별 연구대상자의 중재결과 결측자료(추적관찰 소실 혹은 분석에서 배제)에 대한 영향을 고려한다.	불완전한 중재결과 자료는 비뚤림을 유발할 수 있다. 대부분의 상황에서 가능한 ITT 분석 원칙을 따라야한다(유해효과에는 적절하지 않을 수 있으며, 동등성 시험에도 적합하지 않을 수 있다). 불완전한 중재결과 자료로 인한 비뚤림위험은 코크란 비뚤림위험 도구에서 다룬다. 그러나 통계 분석과 결과에 대한 신중한 해석이 추가적으로 필요하다. 결측치 대치 방법이 고려 될 수 있다.

C71 : 만감도 분석(권고)	
민감도 분석을 통해 주목할만한 가정의 영향, 결측치 대치된 자료, 경계선 의사 결정, 비뚤림위험 높은 연구에 대한 연구결과의 견고성을 평가한다.	결론의 강도가 강화되거나 약화 될 수 있으므로 연구결과가 견고할 때는 주의를 기울여야 한다.

11.15 참고문헌과 읽을거리

- Deeks JJ, Higgins JPT, Altman DG(editors). Chapter 10: Analysing data and undertaking meta-analyses. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA(editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.0(updated July 2019). Cochrane, 2019. Available from www.training.cochrane.org/handbook.

제12장 메타분석 할 수 없는 자료의 제시

핵심 포인트

- 효과 크기의 메타분석은 많은 장점이 있지만, 일차 연구에서 불완전하게 보고된 자료가 있는 상황에서는 다른 합성 방법을 고려할 필요가 있을 수 있다.
- 대체 합성 방법은 필요한 자료의 완전성, 이들이 다루는 가설, 연구 결과에서 도출할 수 있는 결론과 권고사항 등에서 다르다.
- 이러한 방법은 메타분석보다 의료 의사결정에 더 제한된 정보를 제공하지만, 적절한 정당성 없이 일부 연구결과가 다른 연구결과보다 우선시될 수 있는 서술적 설명보다 우월할 수 있다.
- 연구결과에 대한 표와 시각적 표시는 항상 모든 합성과 함께 표시되어야 하며, 메타분석 없는 문헌고찰에서 투명한 보고에 특히 중요하다.
- 합성의 대체물 및 시각 표시 방법은 프로토콜에 명시되어야 한다. 문헌고찰을 작성할 때, 합성 방법의 세부사항을 기술해야 한다.
- 통계적 유의성에 기초한 집계와 관련된 합성방법은 심각한 한계가 있으므로 받아들일 수 없다.

12.1 효과 크기의 메타분석이 불가능한 이유

효과 크기의 메타분석에는 많은 잠재적 장점이 있다. 그러나 메타분석을 수행할 수 없는 상황인 경우 다른 통계 합성 방법을 고려할 수 있다.

메타분석을 할 수 없는 명백한 이유로는 근거의 제한(연구가 없거나 1편인 경우 등), 불완전하게 보고된 중재결과/효과 크기 또는 여러 연구에서 다르게 사용된 효과 크기, 그리고 비뚤림 등을 들 수 있다. 메타분석을 사용하지 않는 다른 일반적인 이유로 너무 많은 임상적 또는 방법론적 다양성 또는 통계적 이질성 때문인 경우가 있지만 메타분석 방법은 적절하게 수행되고 해석될 경우 중요한 통찰력을 제공할 수 있으므로 그러한 경우라도 고려되어야 한다.

12.2 효과 크기의 메타분석이 불가능한 경우 통계적 합성

다양한 통계 합성 방법을 사용할 수 있으며, 이러한 방법은 선호도에 따라 세 가지 범주로 나눌 수 있다(선호되는 것, 받아들일만한 것, 받아들일 수 없는 것). 이 중 선호되는 것은 메타

분석에 의한 통계적 합성이다. 만일 일차 연구에서 불완전하게 보고된 자료로 인해 효과 추정의 메타분석이 불가능할 때 고려될 수 있는 두 가지 방법이 있다. 그것은 '수용할 수 있는 것'과 '수용할 수 없는 것'이다. '수용할 수 있는' 방법은 요구하는 자료의 종류, 가설, 그 사용에 대한 제한, 그리고 도출할 수 있는 결론과 권고사항에 따라 다르다. '수용할 수 없는' 방법은 문제가 되는 부분과 함께 설명한다.

메타분석 방법과 비교하여 '수용 가능' 합성 방법은 의사결정에 제한된 정보만을 제공한다. 그러나 이러한 '수용 가능한' 방법은 연구별 결과 연구를 설명하는 서술보다 우월할 수 있으며, 이는 일부 연구나 연구 결과가 적절한 정당성 없이 다른 연구보다 특권을 갖는 위험성을 수반한다. 또한, 합성이 거의 없거나 전혀 없는 문헌고찰에서 연구 자체를 이해하도록 남겨두는데, 이는 통계적 유의성에 기초한 선정과 같은 간단해 보이지만 문제가 있는 합성 방법을 사용하는 결과를 초래할 수 있다.

모든 방법은 먼저 '표준화된 단위(metric)'의 계산을 수반하며, 그 다음에 합성 방법을 적용한다. 다음 합성 방법 중 하나를 적용할 때, 연구당 하나의 중재 결과만이 합성에 기여하는 것이 중요하다. 메타분석과 유사하게, 통합 결과가 잠재적으로 영향력 있는 의사결정에 대해 확고한지 여부를 검토하기 위해 민감도 분석을 수행할 수 있다.

상세 설명 없이 '서술적 합성' 또는 '서술적 요약'을 수행했다고 명시하지 말고 메타분석 대신 사용되는 구체적인 방법을 보고해야 한다. 선택된 방법의 한계를 기술해야 하며, 사려깊게 결론을 기술해야 한다. 이 세부사항을 보고하는 목적은 합성 과정을 보다 투명하고 재현할 수 있도록 하는 것이고, 적절한 방법과 해석을 보장하는 것이다.

12.2.1 가능한 수용 합성 방법

12.2.1.1 기술통계를 이용한 효과 크기의 요약

분산자료가 없는 경우 메타분석의 가능성이 제한될 수 있지만, 효과는 중앙값, 사분위 범위 및 범위와 같은 기술적 통계를 사용하여 요약할 수 있다. 이러한 통계를 계산하면 '관측된 효과의 범위와 분포는 무엇인가?'라는 질문이 해결된다. 효과를 요약하는 데 사용될 통계를 보고해야 한다. 박스-위스키 또는 버블 그림은 관찰된 효과의 분포를 시각적으로 표시함으로써 요약 통계 보고를 보완한다. 이용 가능한 효과 크기의 표는 효과와 연구를 연결함으로써 독자들에게 투명성을 제공할 것이다.

12.2.1.2 p-값의 결합

기술적 방법: p-값의 결합은 p-값과 효과의 방향을 보고하는 정보 외의 정보가 없거나 최소인 상황에서 고려될 수 있다. p-값을 조합하면 '적어도 하나의 연구에 효과가 있다는 근거가 있는가?'라는 의문이 해결된다. 이용 가능한 방법에는 여러 가지가 있다.

12.2.1.3 효과방향에 따른 집계법

효과의 방향에 기초한 집계 방법은 효과의 방향만 보고되어 있고, 연구 전체에 걸쳐 일관된

효과 측정이나 자료가 보고되지 않는 상황에서 고려될 수 있다. 집계의 본질은 이익을 보여주는 효과의 수와 특정 결과에 위해를 나타내는 효과의 수를 비교하는 것이다. 단, '이익'과 '해'를 정의하는 방법에 차이가 있어 방법의 실시에는 큰 차이가 있다. 주관적 결정이나 통계적 유의성에 기초한 규칙은 문제가 있으므로 피해야 한다.

집계를 적절하게 실시하기 위해, 각 효과 크기는 먼저 관측된 효과 방향에만 근거하여 유익성 또는 위해성을 보여주는 것으로 분류하여 작성한다. 유익성을 나타내는 효과의 개수를 위해성을 나타내는 개수와 비교한다. 통계적 유의성이나 영향의 크기는 분류에서 고려되지 않는다. 부호검정은 '효과에 대한 근거가 있는가?'라는 질문에 대답하는 데 사용될 수 있다.

집계 방법은 문헌고찰의 '자료합성' 부분에 보고해야 한다. 집계에 기반한 합성에 기여하는 연구의 수는 메타분석보다 많을 수 있다. 왜냐하면 집계에 필요한 효과의 방향은 최소가 필요하기 때문이다. 집계의 결과는 종종 허용할 수 없는 집계 방법을 사용하지만 수확도(harvest plot) 및 효과 방향도(effect direction plots)를 도출하는 데 사용된다(섹션 12.3.5 참조). 방법의 한계는 인정되어야 한다.

집계의 방법에 있어서 통계적 유의성에 따른 방법이나 주관적 규칙에 따른 집계를 방법을 사용하여서는 안 된다. 통계적 유의성에 따른 방법은 구체적인 공식에 관계없이, 통계적 중요성에 기초할 때, 모든 것은 심각한 한계를 가지고 있으며 잘못된 결론으로 이어질 수 있다. 주관적 규칙에 따른 집계의 경우, 효과의 방향이나 통계적 유의성 및 효과의 크기 조합을 포함하는 주관적 규칙은 때때로 효과를 분류하는 데 사용된다. 이러한 주관적인 규칙은 효과의 분류에 사용되는 요소, 임계점 및 알고리즘에서 종종 다양하다.

12.3 자료의 시각적 표시 및 표현

자료의 시각적 표시와 표현은 메타분석을 할 수 없는 문헌고찰에서 투명한 보고를 위해 특히 중요하며, 합성이 수행되는지 여부에 관계없이 고려되어야 한다. 표와 그래프 구조 정보는 자료의 패턴을 보여주고 텍스트보다 더 효율적으로 세부 정보를 전달한다. 이것은 해석을 돕고 독자들이 검토 결과의 진실성을 평가하는데 도움을 준다.

12.3.1 연구 전체에서 결과의 구조화된 표

연구 ID에 의해 알파벳순으로 연구를 정렬하는 것은 표에 대한 가장 간단한 접근 방법이지만, 연구를 하위 패널로 그룹화하거나 결과를 해석하는 데 중요한 특성에 의해 정렬할 때 더 많은 정보를 전달할 수 있다. 표에 있는 연구를 그룹화 할 때는 일반적으로 본문에 제시된 합성 구조를 따라야 하며, 이는 검토 질문을 면밀히 반영해야 한다. 이 그룹은 독자들이 발견사항의 기초가 되는 자료를 식별하고 검토 저자의 해석을 검증하는 데 도움이 되어야 한다.

표의 목적이 비교인 경우, 다음 특성 중 하나에 의한 그룹화가 유용할 수 있다.

- (합성 구조에 따라) 문헌고찰 또는 중재결과 영역에서 고려된 비교
- 모집단 하위집단, 세팅 또는 중재 구성요소를 포함한 잠재적 효과 변경자와 같은 자료 패턴을 나타낼 수 있는 연구 특성

표의 목적이 완전하고 투명한 자료 보고인 경우, 연구를 정렬하여 가장 중요한 자료의 중요도를 증가시킨다. 타당하고 신뢰할 수 있는 근거를 고려해야 한다. 가능성에는 다음이 포함된다.

- 근거의 확실성.
- 비뚤림위험, 연구 크기 또는 연구설계 특성
- 연구가 문헌고찰 문제를 직접적으로 다루는 방법, 예들 들어 중재결과의 적절성과 타당성

연구의 특성에 의한 그룹화의 한 가지 단점은 예를 들어 본문 내용과 표간의 상호 참조와 같이 연구ID에 의해 표가 정렬되는 경우보다 특정 연구를 찾는 것이 더 어려울 수 있다는 것이다. 범주 내에서 연구ID에 의한 정렬하여 부분적으로 이를 해결할 수 있다.

12.3.2 도표를 이용한 제시

12.3.2.1 숲 그림

메타분석을 수행하지 않을 때는 연구 결과를 표시하는 것이 중요하다. 숲 그림은 개별 연구 결과의 해석에 도움이 될 수 있으며, 특히 결과 해석에 중요한 특성(예: 용량 및 효과 크기, 표본 크기)에 의해 연구가 정렬될 때 자료의 전반적인 패턴을 전달할 수 있다. 마찬가지로, 인구집단의 하위군, 중재의 변이 또는 비뚤림의 위험과 같이 효과변경인자로 간주되는 특성에 기반한 하위 군의 연구 그룹화는 연구 전체의 차이를 탐색하고 설명하는 데 도움이 될 수 있다.

12.3.2.2 박스-위스키 그림(상자그림) 및 버블 그림

상자그림은 효과 크기의 분포를 시각적으로 보여준다. 상자그림은 관례적으로 다섯 가지 값을 제시한다. 상자의 상한과 하한은 각각 75번째와 25번째 백분위수를 나타낸다. 상자 안의 선은 50번째 백분위수(중간)를 나타내며, 수염은 극단 값을 나타낸다. 효과 크기의 분포를 시각적으로 비교할 수 있도록 다중 상자 그림을 병렬로 제시할 수 있다. 복수의 상자그림을 표시할 때 상자 너비는 각각에 기여하는 연구의 수를 나타내도록 변경할 수 있다. 그림의 공통적 용도는 독자가 신속하고 정확하게 해석하도록 도와준다는 것이다. 그러나 그림에 기여하는 개별 연구는 식별되지 않으며, 연구가 거의 없을 때는 그림이 적절하지 않다.

버블그림을 사용하여 효과 분포를 시각적으로 표시할 수 있으며, 연구가 거의 없을 때 상자그림보다 더 적합하다. 버블그림은 버블의 위치, 크기, 색상을 통해 여러 값을 표시할 수 있는 산점 그림이다.

12.3.2.3 앨버트로스 그림

앨버트로스 그림은 연구 내에서 결과 보고가 최소인 경우 기본 중재 효과 크기를 대략적으로 제시할 수 있다. 그림에는 각 결과에 대해 표본 크기와 효과 방향(또는 동일하게 단측 p-값과 표본 크기), 그리고 양면 p-값만 필요하다. 그림은 양면 p-값에 대한 연구 표본 크기의 산점도인데, 여기서 결과는 효과 방향으로 구분된다. 그림에 겹쳐진 것은 '효과 크기 등고선'(그림의 이름 삽입)이다. 이러한 등고선은 p-값을 계산하는 데 사용되는 자료 유형(예: 연속, 이진)과 통계적 방법에 한정된다. 등고선은 연구의 대략적인 효과 크기를 해석할 수 있으며, 그렇지 않으면 결과의 제한된 보고로 인해 불가능할 것이다.

12.3.2.4 하베스트 그림 및 효과 방향도

하베스트 그림은 집계 결과를 시각적으로 보여준다. 그림에서 효과의 분류에 기초한 연구가 함께 분류된다. 각 연구는 분류에 따라 배치된 막대로 표현된다. 막대는 '시각적 가중치'(높이 또는 너비에 따라)가 될 수 있으며 연구 및 예후 특성을 강조하기 위해 주석을 달 수 있다(예: 비뚤림위험 영역, 연구설계, 표본 크기). 주석을 사용하여 연구를 식별할 수도 있다. 일련의 그림은 예를 들어 다른 중재 또는 결과 도메인의 집계 결과를 표시하는 매트릭스에 결합될 수 있다.

효과 방향도는 효과 방향에 대한 정보를 모두 표시한다는 점에서 하베스트 그림과 개념적으로 유사하다. 효과 방향도의 첫 번째 버전에서는 단일 연구 내의 각 결과에 대한 효과 방향이 표시되고, 두 번째 버전은 여러 연구의 결과 영역에 대한 효과 방향을 표시한다. 이 두 번째 버전에서는 한 영역 내의 모든 결과에 대한 효과 방향(예: 예후 영역 '호흡'에서 '수면 방해', '호흡 제한 음성', '운동 중 졸음')을 '합성'하기 위해 알고리즘을 먼저 적용한다. 이 알고리즘은 일관된 방향과 통계적 유의성에 있는 영향의 비율에 기초한다.

12.4 참고문헌과 읽을거리

- McKenzie JE, Brennan SE. Chapter 12: Synthesizing and presenting findings using other methods. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA(editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.0(updated July 2019). Cochrane, 2019. Available from www.training.cochrane.org/handbook.

제13장 네트워크 메타분석

핵심 포인트

- 네트워크 메타분석(Network Meta Analysis)은 선정된 연구 네트워크에 안에서 직접적 근거와 간접적 근거를 모두 통합하여 단일 분석에서 3개 이상의 중재를 동시에 비교하는 기법이다.
- 네트워크 메타분석은 네트워크에서 어떤 중재 사이의 상대적 효과에 대한 추정치를 산출하며, 일반적으로 단일 직접 또는 간접 추정치보다 더 정밀한 추정치를 산출한다. 또한 중재의 순위 및 계층의 추정을 할 수 있게 한다.
- 유효한 네트워크 메타분석은 분석에 포함된 여러 연구들이 평균적으로 상대적 효과에 영향을 미칠 수 있는 모든 중요한 요소에서 유사하다는 가정에 의존한다.
- 비일관성은 특정 중재 비교에 대한 다른 정보 출처(예: 직접 및 간접)가 일치하지 않을 때 발생한다.
- 네트워크 메타분석에서 근거에 대한 신뢰도를 평가하는 것은 각 직접 비교에 대한 신뢰도를 평가하는 것으로 시작한다. 도메인별 평가를 결합하여 근거에 대한 전반적인 신뢰도를 결정한다.

13.1 네트워크 메타분석이란?

특정하게 주어진 조건 하에서는 몇 개 또는 아주 많은 경쟁적인 중재가 있을 수도 있다. 대안적 중재 중 하나를 결정해야 하는 사람들은 모든 관련 중재를 포함하는 단일 검토가 유익할 수 있으며, 이들의 비교 효과와 위해성의 가능성을 제시할 수 있다. 네트워크 메타분석은 그러한 검토를 위한 분석 옵션을 제공한다.

직접 비교를 통해 세 개 이상의 중재를 연결하는 모든 일련의 연구는 중재 네트워크를 형성한다. 중재의 네트워크에서는 중재들 사이를 간접 비교하는 여러 가지 방법이 있을 수 있다. 이는 연구 내에서 직접 수행되지 않은 비교로서, 이용 가능한 직접 중재 효과 크기의 수학적 조합을 사용하여 추정할 수 있다. 네트워크 메타분석은 단일 분석에서 중재 네트워크를 통해 직접 및 간접 추정치를 결합한다. 동의어 용어로 혼합 비교와 다중 메타분석이 있다.

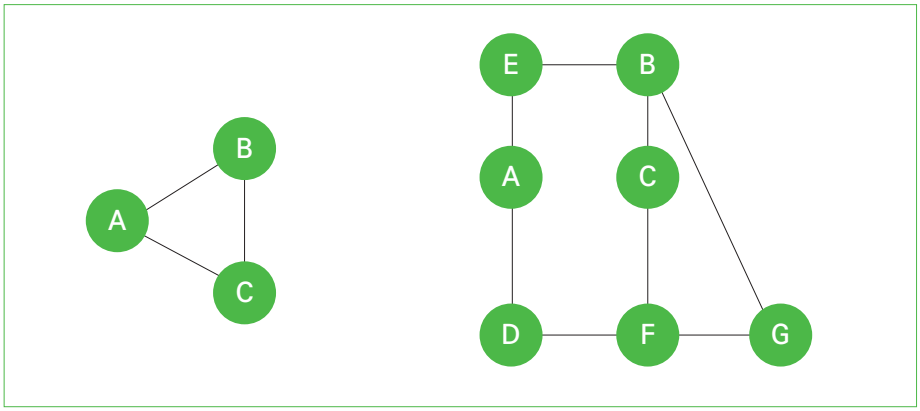


그림 13-1. 네트워크메타분석의 비교

네트워크 메타분석의 장점은 이용 가능한 모든 직접 및 간접 근거를 이용할 수 있다는 것이다. 또한 단일 직접 또는 간접 추정치와 비교하여 중재 효과에 대한 더 정밀한 추정치를 산출할 수 있으며, 동일한 분석에 대한 모든 관심 중재의 동시 비교를 통해 특정 중재결과에 대한 상대 순위를 추정할 수 있다.

13.2 네트워크 메타분석을 위한 중요개념

네트워크 메타분석 방법론의 핵심은 간접 비교의 개념이다. 직접 비교한 연구가 없을 때 두 중재의 상대적 효과를 추정하기 위해 간접 비교가 필요하다.

13.2.1 간접비교

간접 비교는 임상시험 내에서 직접 비교되지 않은 두 가지 중재의 상대적 효과를 추정할 수 있게 해준다. 예를 들어 아래 그림에서 중재 A와 B를 직접 비교하는 무작위배정 임상시험과 A와 C를 비교하는 무작위배정 임상시험이 있다고 가정하자. 또한 이러한 무작위배정 임상시험들이 표준화 되어 있다고 가정하자, A대B('AB')와 A대C('AC')에 대한 중재 효과의 직접적인 추정치를 도출하기 위해 쌍으로 메타분석을 실시했다. 아래그림에서 실선은 사용 가능한 근거를 나타낸다. 중재 C에 대한 B의 효과에 대해 알고자한다. 그리고 그 점선은 직접적인 근거가 없는 이 비교를 묘사한다.

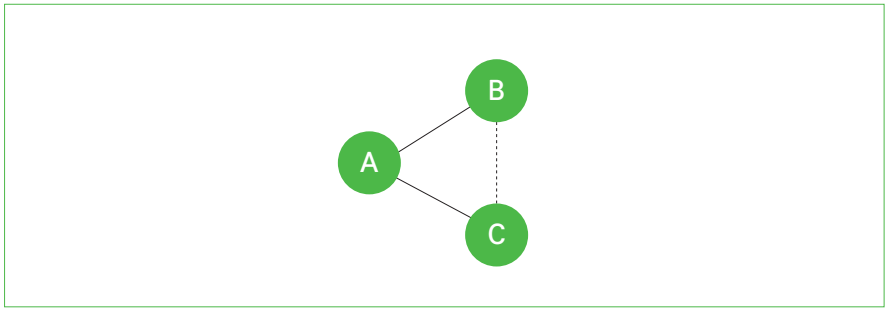


그림 13-2. 간접비교

간접 비교를 이해하는 한 가지 방법은 BC 비교(B 대 C)를 C에 대한 B의 이득(B/C)을 나타내는 것으로 생각하는 것이다. 다른 모든 것이 같다면, C에 대한 B의 이득은 A에 대한 B의 이득과 B에 대한 A의 이득과 동일하다.따라서 예를 들어, B에 대한 C이익을 기술하는 간접 비교는 A에 대한 B의 이익과 C에 대한 A의 이익의 합으로 생각할 수 있다. 이합을 수학적으로 표현하면

$$\text{indirect}\ (B vs\, C) = \text{direct}\,(B vs\, A) + \text{direct}\ (A vs\, C)$$

이를 차의 형태로 표현하기도 한다.

$$\text{indirect}\ (B\, vs\, C) = \text{direct}\ (A\, vs\, C) - \text{direct}\ (A\, vs\, B)$$

직접 C 대 A와 B 대 A의 메타분석에서 중재 효과의 요약 통계 간의 차이가 중재효과에 대한 간접적인 추정치를 제공하도록 한다.

두 가지 직접 비교(세 가지 중재)가 있는 이 단순한 사례의 경우, 표준 메타분석 프로그램(Revman 포함)을 사용하여 부분군 분석을 수행하여 분석을 수행할 수 있다. 두 가지 직접 비교를 다루는 연구(즉, A 대 B 및 A 대 C)는 메타분석에서 두 개의 부분군으로 취급할 수 있다. 두 부분군의 요약 효과 차이는 간접 비교에 대한 추정치가 된다. 대부분의 소프트웨어는 간접 효과 크기의 추정 분산에 기초한 부분군 간 차이의 통계적 유의성에 대한 p-값을 제공한다.

$$\begin{aligned}&\text{Variance}[\text{indirect}\ (B vs\, C)] \\ &= \text{Varaince}[\text{direct}\ (A\, vs\, C)] + \text{Variance}[\text{direct}\ (\ A\, vs\, B)]\end{aligned}$$

여기서 $\text{Varaince}[\text{direct}\ (A\, vs\, C)]$와 $\text{Variance}[\text{direct}\ (\ A\, vs\, B)]$는 각각의 직접 추정치의 분산이다.

간접 요약 효과에 대한 95% 신뢰 구간은 다음 공식에 의해 생성된다.

$$\mathrm{indirect}(B\,vs\,C) \pm 1.96 \times \sqrt{\mathrm{Variance}[\mathrm{indirect}\ \ (B\,vs\,C)]}$$

이 방법은 무작위배정 임상시험의 각 군의 중재 효과를 사용하므로 시험 내 임의 배정을 유지한다. 만일 연구 전체에 걸쳐 단일 군을 통합(pooling)하고(예: 모든 B 군과 모든 C군) 합쳐진 B군과 C군(즉, 자료를 하나의 큰 무작위배정 실험에서 나온 것처럼 취급)을 직접 비교했다면, 분석은 임상 무작위배정의 장점을 잃게 될 것이다. 그런 접근법을 사용해서는 안 된다.

13.2.2 이행성(transitivity)

간접 비교의 기본 가정은 치료 A를 통해 B 대 C의 진정한 상대적 효과를 알 수 있다는 것이다. 이 관계는 수학적으로 다음과 같이 표현할 수 있다.

$$\mathrm{effect\ of\ B\ versus\ C} = (\mathrm{effect\ of\ A\ versus\ C}) - (\mathrm{effect\ of\ A\ versus\ B})$$

즉, 중재 A(그림13-2)를 통해 중재 B와 C를 비교할 수 있다는 것을 의미한다.

간접 비교는 무작위배정 임상시험 전반에 걸친 관찰된 근거를 제공하며 교란 요인과 같은 관찰 연구의 비뚤림이 존재할 수 있다. 간접 비교의 유효성은 다양한 무작위배정 임상시험에서 수행되는 중재 비교 이외의 모든 중요한 요소가 평균적으로 유사하다는 가정에서 유효하다.

이러한 요구사항을 설명하기 위해 이행성(transitivity)라는 용어를 사용한다. 그것은 일관성이라는 통계적 개념과 밀접하게 관련되어 있다. 두 용어의 관계는 짝 메타분석에서 다양성(diversity)과 이질성(heterogeneity) 사이의 관계와 유사하다.

서로 다른 중재를 비교하는 연구는 특성이 다양할 수 있다. 때때로 이러한 특성은 중재의 효과와 연관된다. 그러한 특징들을 효과 변경인자라고 부른다. A 대 B와 A 대 C의 무작위배정 임상시험에서 효과 변경인자가 다른 경우, 간접 비교하는 것은 적절하지 않을 것이다.

이행성은 두 가지 상대적 효과에 영향을 미칠 수 있는 특성(효과 변경인자)가 A 대 B 연구와 A 대 C 연구에서 중재 A가 나타날 때 유사해야 한다는 것이다.

이행성은 모든 경쟁 중재의 체계적 문헌고찰에서 공동으로 임의 배정할 수 있어야 한다. 즉, 모든 중재는 단일 다중군 무작위배정 임상시험에서 동시에 비교되는 것을 상상할 수 있다. 이것을 보는 또 다른 방법은, 어떤 특정한 시험에서, '결측' 중재(시험에 포함되지 않은 중재)가 효과와 무관한 이유로 결측된 것으로 간주될 수 있다는 것이다.

13.2.3 간접 비교 및 네트워크 메타분석의 타당성

13.2.3.1 직접 및 간접 근거 결합

중재에 대한 직접적인 비교와 간접적인 비교가 모두 존재하는 경우가 있다. 주요 이행성 가정이 합리적이라고 간주되는 경우, 직접 및 간접 추정을 동시에 고려해야 한다. 직접 및 간접 중재 효과를 특정 비교에 사용할 수 있는 경우, 이러한 효과를 단일 효과 크기로 합성할 수 있다. 이 요약 효과는 때때로 중재 효과의 통합 또는 혼합 추정치라고 불린다. 혼합 추정치는 직접 및 간접 요약 추정치의 역분산 가중평균으로 계산할 수 있다.

혼합 추정치는 간접 비교를 포함하므로, 이 추정치는 이행성 가정에 의존한다. 이행성의 위반은 간접 및 결합 추정치의 유효성을 위협한다.

13.2.3.2 일관성(또는 consistency)

이행성의 중요한 가정은 다른 비교간의 임상적, 방법론적 잠재적인 변이와 관련이 있다. 이러한 차이 때문에 다른 근거 출처 간의 추정이 불일치한 형태로 반영될 수 있다. 이행성의 통계적 표현을 일반적으로 일관성(coherence 또는 consistency)라고 부른다. 우리는 표준 메타분석 내에서 개념적으로 불일치(또는 이질성)를 구별하기 위해 전자(일관성)를 사용할 것이다. 일관성은 다른 근거의 출처(직접과 간접)가 서로 일치한다는 것을 암시한다.

일관성 가정은 특정 비교에 대한 진정한 직접 및 간접 중재 효과가 동일하다고 기술하는 일관성 방정식에 의해 수학적으로 표현된다.

$$true\ (B\ vs\ C) = true\ (A\ vs\ C) - true\ (A\ vs\ B)$$

13.3 결과의 합성

13.3.1 네트워크 메타분석은 무엇을 추정하는가?

연결된 네트워크에서 일관성 방정식은 중재 효과 사이의 수학적 연계를 제공하므로, 일부 효과는 다른 이행성 가정을 사용하여 계산할 수 있다. 이것은 모든 쌍의 비교가 독립적으로 추정되지 않는다는 것을 의미한다. 사실, 네트워크 메타분석에서 추정되어야 하는 비교의 수는 중재의 수를 뺀 것과 같다.

네트워크 메타분석의 주요 결과는 모든 기본적인 비교에 대한 중재 효과에 대한 네트워크 추정치의 집합이다. 일관성 방정식을 사용하여 분석 후 다른 비교에 대한 추정치를 구한다. 어떤 비교를 기본 비교로 선택하느냐는 중요하지 않다. 종종 하나의 중재를 참조로 식별하고, 기본 비교를 이 참조에 대한 각 다른 중재의 효과로 정의한다.

13.3.2 메타 회귀 분석을 이용한 직접 및 간접 근거 합성

네트워크 메타분석은 몇 가지 접근법을 사용하여 수행할 수 있다. 모든 접근방식에 대한 주요 기술적 요건은 분석에 포함된 모든 중재가 '연결된' 네트워크를 형성한다는 것이다. 많은 네트워크에 사용되는 간단한 접근방식은 메타 회귀 분석을 사용하는 것이다. 이 접근방 식은 네트워크에 다중 군 임상시험이 없는 한 작동한다.

연구 하위군 간의 차이도 메타 회귀 분석을 통해 조사할 수 있다. 표준 메타 회귀 분석을 사용하여 단일 간접 비교를 수행하는 경우, 단일 더미 변수를 사용하여 각 연구의 결과가 어떤 직접 비교와 관련되는지 아니면 다른 비교와 관련되는지 여부를 지정한다(더미 변수는 연구에서 어떤 비교가 이루어지는지 나타내기 위해 1 또는 0으로 코드화됨).

13.3.3 네트워크 메타분석 수행

다중 군 시험이 있을 때 사용할 수 있는 네트워크 메타분석을 위해 특별히 설계된 접근법을 소개한다. 네트워크 메타분석을 수행하는 일반적인 접근방식은 Bayesian 프레임워크 내에서 일반적으로 구현되는 계층적 모형을 사용하는 것이다. Bayesian 접근법을 위한 소프트웨어 옵션에는 WinBUGS와 OpenBUGs 등이 있다.

초기에 복수의 결과를 공동으로 합성하기 위해 개발된 다중 메타분석 방법은 네트워크 메타분석 수행에 대한 대안적 접근법을 제공한다. 다중 메타분석 접근법은 기본 비교 집합에 대한 분석(예: 공통 참조 중재에 대한 각 중재)에 초점을 맞추고 이를 다른 결과와 유사한 것으로 취급한다. 한 연구는 하나 이상의 기본 비교에 대해 보고할 수 있다. 예를 들어, 세 군 무작위배정 임상시험에는 두 개의 비교가 있다. 기본 비교를 대상으로 하지 않는 연구(예: 공통 참조 중재가 포함되지 않은 연구)의 경우 자료 증강이라고 알려진 기법을 사용하여 적절한 매개변수를 지정할 수 있다. 이 방법은 Stata에서 이용할 수 있는 네트워크 매크로에서 구현된다.

13.3.3.1 이질성에 대한 가정

이질성은 동일한 중재 쌍을 직접 비교하는 무작위배정 임상시험 사이의 근본적인 차이를 반영한다. 쌍으로 이루어진 메타분석에서 중요한 이질성의 존재는 요약 효과의 해석을 어렵게 할 수 있다. 네트워크 메타분석 추정치는 직접 비교와 간접 비교를 통해 이용 가능한 직접 추정치의 조합이므로, 한 가지 비교에 대한 연구 사이의 이질성은 많은 다른 비교의 결과에 영향을 미칠 수 있다.

네트워크의 모든 비교에 대해 이질성의 양이 동일하다고 가정하는 것이 일반적이다. 이것은 비교 고유의 이질성을 가정하는 것에 비해 세 가지 장점이 있다. 첫째, 비교 전반에 걸쳐 정보를 공유하여 한두 번의 시험과의 비교만으로 여러 시험과의 비교에서 이질성에 대한 정보를 차용할 수 있다. 둘째로, 더 많은 자료가 추정치에 기여하기 때문에 이질성은 더 정확하게 추정되며, 그 결과 일반적으로 중재 효과의 더 정확한 추정치가 발생한다. 셋째, 공통 이질성을

가정하면 비교 고유의 이질성을 가정하는 것보다 모형 추정을 계산적으로 쉽게 할 수 있다.

이질성 가정의 선택은 통계적 특성과 더불어 자료에 대한 임상적, 방법론적 이해와 가정 신뢰성 평가에 기초해야 한다.

13.3.3.2 중재의 순위

네트워크 메타분석의 특징 중 하나는 특정 결과에 대한 경쟁 중재의 상대 순위를 추정할 수 있다는 것이다. 확률 순위를 매기는 것, 네트워크의 다른 중재와 비교할 때 중재가 특정 순위(첫 번째, 두 번째 등)에 있을 확률을 자주 사용된다. 순위결정 확률은 결과에 따라 달라질 수 있다. 추정 수량의 경우, 순위 확률은 일부 변동성과 함께 추정한다. 따라서 변동성을 고려하지 않고 최고 순위에 오를 확률에만 기초한 추론은 오해를 불러일으키므로 피해야 한다.

평균 순위, 중위수 및 누적 순위 확률과 같은 순위 측정은 가능한 모든 순위에 대한 추정 확률을 요약하고 상대 순위의 불확실성을 설명한다.

13.3.4 근거 출처 간 불일치(비일관성)

13.3.4.1 비일관성이란?

특정 상대적 효과에 대한 다른 정보 출처가 불일치할 때 비일관성이 발생한다. 네트워크 메타분석에 관한 많은 문헌에서 비일관성(inconsistency)라는 용어가 비일관성(incoherence) 대신 사용된다.

네트워크 그래프에서 비일관성 크기는 모든 쌍방향 비교에서 직간접 요약 추정치 사이의 절대 차이로 측정될 수 있다. 비일치성을 검출하는 이 방법을 'loop-specific approach'이라고 한다. 얻어진 통계는 대개 비일관성인자(incoherence factor) 또는 inconsistency factor(IF)라고 한다. 비일관성은 다음과 같이 추정할 수 있다.

$$\mathrm{IF} = \left|\mathrm{direct}\left(\mathrm{B}\ vs\ \mathrm{C}\right) - \mathrm{indirect}\left(\mathrm{B}\ vs\ C\right)\right|$$

IF는 직접 효과 크기와 간접 효과 크기 사이의 불일치 수준을 측정한다.

여러 루프를 사용하는 중재 네트워크에서 비일관성을 평가하기 위한 몇 가지 접근법이 제안되었다. 지역(local) 접근방식은 네트워크 영역을 별도로 평가하여 가능한 '비일관성 지점'을 탐지하는 반면, 전체(global) 접근방식은 전체 네트워크에서 일관성을 평가한다.

13.3.4.2 지역 비일관성 평가 접근법

비일관성을 조사하기 위한 지역 접근방식은 SIDE(Separating Indirect from Direct Evidence, 직접 근거에서 간접 분리)이다. 이것은(사용 가능한 경우) 직접 추정치와 간접 추정치를 비교하여 네트워크의 모든 쌍별 비교에 대한 IF를 전체 네트워크에서 추정한다.

비일관성에 대한 검정은 검정력이 낮으므로 존재하는 경우에도 통계적으로 유의한 비일관성을 감지하지 못할 수 있다. 이것은 통계적으로 유의한 비일관성이 없다는 것이 비일관성이 없다는 근거가 되지 못함을 의미한다. 비일관성 요인에 대한 신뢰 구간을 고려해야 하며, 직접 근거와 간접 근거 간의 임상적으로 중요한 비일관성을 제안할 만큼 충분히 큰 값을 포함하는지 여부를 결정해야 한다.

13.3.4.3 전체 비일관성 평가 접근법

네트워크에서 전체의 비일관성은 비일관성 모형을 통해 평가되고 감지될 수 있다. 일관성 방정식을 완화시키고, 직간접적으로 추정할 때 중재 효과가 달라질 수 있도록 한다. 분명히 복잡한 네트워크에는 여러 개의 IF가 있을 것이다. 네트워크가 일관성을 유지하려면 모든 IF가 0에 가까워야 한다. 이는 매크로의 Stata에서 사용할 수 있는 통계 검정를 통해 정식으로 검정할 수 있다. 이 모형의 확장은 다른 일련의 중재가 관련된 연구에서 효과 크기가 측정될 때 비일관성을 측정하는 경우 제안되었다.

Q-검정 및 I^2통계와 같은 측정은 쌍 메타분석에서 이질성 평가에 일반적으로 사용되며, 네트워크 메타분석에서 이질성 및 비일관성에 대한 평가를 위해 개발되었으며, 이는 R의 패키지 netmeta에서 구현되었다.

13.3.4.4 비일관성에 대한 결론 도출

지역적 접근법과 전체적 접근법을 모두 사용하고 그들의 결과를 공동으로 검토하여 비일관성에 대한 추론을 할 것을 제안한다. 표준 쌍 메타분석에서 통계적 이질성에 대한 검정의 경우, 비일관성을 검출하기 위한 검정은 시뮬레이션과 경험적 연구에서 보듯이 존재하는 경우 비일관성에 대한 검정력이 부족한 경우가 많다.

결론도 단순히 통계적 유의성을 고려한 것이 아니라 비일관성 요인의 신뢰구간에 포함된 값의 범위를 해석하여 도출해야 한다. 연구자들은 통계적으로 유의한 비일관성이 없다고 해서 네트워크의 일관성이 보장되는 것은 아니며, 이는 분석을 수행하기 전에 항상 효과 변경인자를 검사하여 평가해야 한다는 것을 기억해야 한다.

비일관성이 감지되면 가능한 설명을 하여야 한다. 자료 수집의 오류, 광범위한 적격성 기준 및 효과 변경인자의 불균형 분포가 비일관성을 초래했을 수 있다.

13.4 네트워크 메타분석 결과에 대한 신뢰도 평가

각 쌍의 중재적 비교에 대한 근거의 신뢰도를 평가하기 위해 GRADE 접근방식을 사용하는 것이 좋다. 이 접근법은 특정 쌍 비교 무작위배정 임상시험에 대한 근거에 높은 신뢰도를 가정한 다음, 연구 제한, 비직접성, 비일관성, 비정밀성 및 출판 비뚤림의 다섯 가지 요인에 의해 근거의 등급을 낮추도록 한다.

중재 네트워크의 근거에 대한 신뢰도를 평가하는 것은 쌍 메타분석보다 더 어렵다.

13.5 네트워크 메타분석 제시

네트워크 메타분석의 결과를 보고할 때 체계적 문헌고찰의 보고를 위한 PRISMA 확장 선언문을 고려해야 한다. 주요 그래픽 및 수치 요약에는 네트워크 그림, 치료제 사이의 상대적 효과에 대한 표, 이질성과 비일관성의 측정이 포함된다.

13.5.1 네트워크 메타분석의 근거 기반 제시

네트워크 도표(network diagram)는 네트워크의 구조를 설명하는 편리한 방법을 제공한다. 연구 수준 또는 비교 수준 특성에 대한 정보를 포함하도록 수정될 수 있다. 예를 들어 선의 두께는 직접 비교에 선정된 연구 또는 환자의 수 또는 잠재적 효과 변경인자의 비교의 특정 평균을 반영할 수 있다. 후자를 사용하여, 네트워크 도표를 네트워크에서의 이행성 평가를 위한 첫 번째 단계로 간주할 수 있다. 그림 13-3의 예에서 연구대상자의 나이는 잠재적 효과변경인자로 간주되어 왔다. 선의 두께는 B 대 D와 C 대 D 비교 내 특정 인자의 비교가 다른 두 가지 직접 비교와 상당히 다른 것처럼 보인다는 것을 의미한다.

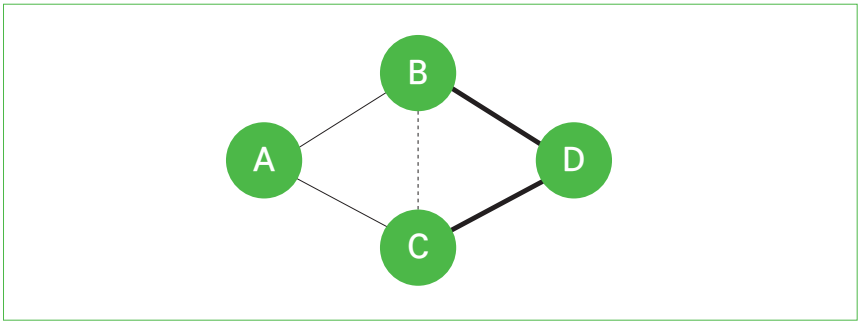


그림 13-3. 네트워크 도표

네트워크에 설계 제한이 있는 연구를 포함(예: 눈가림법, 부적절한 배정 은폐)하는 것은 종종 조사 결과의 타당성을 위협한다. 중재 네트워크에서 색상 선을 사용하면 특정 직접 비교에서 그러한 연구의 존재를 나타낼 수 있다.

13.5.2 결과 표시

쌍으로 이루어진 메타분석과는 달리, 네트워크 메타분석의 결과는 표준 숲 그림과 같은 하나의 그림으로 쉽게 요약할 수 없다. 특히 많은 비교를 수반하는 많은 경쟁적 중재가 있는 네트워크의 경우, 결과를 간결하고 이해할 수 있는 방식으로 제시하는 것은 어렵다.

모든 중재 쌍에 대한 중재 효과의 요약 통계는 네트워크 메타분석에서 가장 중요한 출력물

이다. 일부 비교 결과는 공간 제한으로 인해 제시되며, 보고될 조사 결과의 선택은 연구 문제와 독자에 따라 다를 수 있다. 이 경우, 모든 결과를 상세히 나타내기 위해 추가 수치와 표를 사용할 필요가 있다. 또한 결과의 간결한 요약을 제공하고 의사결정을 용이하게 할 수 있는 보완적 산출물로 중재안의 상대적 순위를 보고하기를 원할 수 있다. 이를 위해 상대적 효과와 상대적 순위를 모두 공동 표시할 것을 권장한다.

13.5.2.1 중재 효과의 표시 및 순위

신뢰 등급은 유용한 형식이다. 또한 네트워크 메타분석 결과의 표시를 위해 다양한 그래픽 도구가 제안되었다. 신뢰 구간과의 쌍방향 비교에 대한 요약 상대적 효과는 숲 그림에 표시할 수 있다.

모든 가능한 순위에 대한 순위결정 확률은 평점그램으로 알려진 확률선을 그려서 제시할 수 있으며, 각 중재에 대한 순위결정 확률의 분포를 보여준다. 2개(경쟁) 결과의 상대 랭킹은 2차원 산점도에서 공동으로 표시할 수 있다. 네트워크 메타분석의 공동 상대 효과와 상대 순위를 제시하는 다양한 방법에 대한 확장된 설명에서 이용할 수 있다.

13.5.2.2 이질성과 비일관성의 표시

중재 네트워크의 이질성 수준은 일반적으로 네트워크의 모든 비교에서 공통으로 가정되는 연구간 분산 의 크기를 통해 표현할 수 있다. 추정 가 이질성의 존재를 시사하는지 여부에 대한 판단은 비교되는 중재안의 유형과 임상 결과에 따라 달라진다. 특정 임상 환경에 특정한 의 추정값에 대한 더 많은 확장된 논의를 이용할 수 있다. 네트워크에서 추정된 모든 비일관성 인자와 그 불확실성을 나타내는 숲 그림은 국지적 비일관성을 표시하기 위해 사용할 수 있다. 전체적 비일관성 평가 결과는 통계 비일관성 검정의 p값과 비일관성 대한 통계량으로 요약할 수 있다.

13.5.2.3 '결과 요약' 표

'결과 요약' 표의 목적은 이용 가능한 자료, 근거에 대한 신뢰 정도 및 중재 효과의 측면에서 핵심 정보를 간결하게 제공하는 것이다. 표를 제공하는 것은 여러 중재들을 동시에 비교하는 검토에서 더 어려운데, 이 검토에서는 중재 쌍들 간에 매우 많은 수의 비교가 수반된다. 일반적인 원칙은 다중 중재의 비교가 네트워크 메타분석의 주요 특징이기 때문에 '결과 요약' 표의 구조를 구성하기 쉽다. 이것은 쌍으로 비교하기 위한 '결과 요약' 표와 대조되는데, 이 표의 주된 강점은 다른 결과에 대한 효과의 비교를 용이하게 하는 것이다. 그럼에도 불구하고, 네트워크 메타분석 결과를 서로 다른 결과 간에 비교할 수 있는 것은 여전히 중요하다. 이것은 2차원에서 거의 해결이 불가능한 과제를 제공한다. 한 가지 잠재적인 해결책은 사용자가 중재 간에 비교를 강조할지 아니면 결과 간에 비교를 강조할지를 선택할 수 있는 대화형 표시장치다.

작은 중재 네트워크(5개 미만)의 경우, 각 주요 결과에 대해 별도의 '결과 요약' 표가 작성될 수 있다. 그러나 많은(5개 이상의) 경쟁 중재가 있는 경우, 연구자들은 일반적으로 쌍으로 구성된 비교의 수를 줄이고 보고할 필요가 있다. 문헌고찰 저자는 '결과 요약' 표에 보고된 비교의 선택에 대한 명확한 근거를 제시해야 한다. 예를 들어, 그들은 중재의 결정 집합에 해당하는 쌍의 비교만 포함할 수 있다. 즉, 결론을 도출하는 직접적인 관심사인 중재군이다. 의사결

정 집합과 더 넓은 종합 대조 집합(분석에 포함된 모든 중재)의 구별은 문헌고찰의 프로토콜에 명시되어져야 한다. 의사결정 세트가 여전히 너무 큰 경우, 연구자들은 임상실무에 중요도를 가정하여 '결과 요약' 표의 비교를 선택할 수 있을 것이다. 예를 들어, 가장 일반적으로 사용되는 세 가지 또는 네 가지 가장 효과적인 중재 간의 비교를 가장 많이사용되는 중재를 대조군으로 하여 보고할 수 있다.

13.6 결론

네트워크 메타분석은 복수의 중재의 비교 효과를 알 수 있는 방법이지만, 표준 메타분석보다 통계적으로 복잡하기 때문에 이 방법을 사용하는데는 주의가 필요하다. 또한 네트워크 메타분석은 일반적으로 광범위한 연구 질문을 하기 때문에, 일반적으로 표준 메타분석보다 선별부터 분석까지 체계적인 문헌고찰의 각 단계에서 더 많은 연구가 필요하다. 따라서 시작 전에 필요한 전문성, 시간 및 자원을 예측하는 것이 중요하다.

유효한 간접 비교와 네트워크 메타분석은 일관성 있는 근거 기반을 필요로 한다. 연구 질문의 공식화 및 적격성 기준을 결정할 때, 이행성의 가정과 관련된 모집단과 중재가 고려될 필요가 있다. 네트워크 메타분석은 서로 다른 일련의 중재들을 비교하는 연구가 결합될 수 있을 정도로 유사할 때만 유효하다. 적절하게 수행되면 단일 직접 또는 간접 추정치보다 더 정확한 상대적 효과 크기를 제공한다. 네트워크 메타분석은 서로 직접 비교한 적이 없는 것을 포함하여 어떤 쌍의 중재 사이에 추정치를 산출할 수 있다. 네트워크 메타분석은 또한 중재의 순위 및 계층의 추정을 허용한다. 결과를 해석하고 네트워크 메타분석에서 결론을 도출할 때, 특히 불협화음이나 기타 잠재적 비뚤림이 있는 경우 많은 주의를 기울여야 한다.

13.7 참고문헌과 읽을거리

- Chaimani A, Caldwell DM, Li T, Higgins JPT, Salanti G. Chapter 11: Undertaking network meta-analyses. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA(editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.0(updated July 2019). Cochrane, 2019. Available from www.training.cochrane.org/handbook.

제14장 자료분석과 결과제시

핵심 포인트

- 이 장에서는 문헌고찰 결과를 효과적으로 전달하기 위해 합성결과를 해석하는 지침을 제공한다.
- 필요 치료 수(NNT)를 포함하여, 이분형 중재결과 자료에 대한 상대 및 절대 효과를 계산, 제시 및 해석하기 위한 방법이 제시된다.
- 연속형 중재결과 측정을 위해 자연 단위를 사용하여 연구에 대한 요약 결과를 제공하거나 모든 연구가 동일한 척도를 사용할 때 최소 중요 차이(MID:minimal important differences)로 제시할 수 있다. 연구에서 동일한 구성을 다른 척도를 측정하였다면 표준화된 평균 차이(SMD: standardized mean difference)를 제시하거나 평균비와 같은 대안적인 효과 크기를 사용해야 한다.
- 합성결과를 '통계적으로 유의함', '통계적으로 유의하지 않음' 또는 '중요하지 않음'으로 설명하거나 p값의 임계치에 과도하게 의존해서는 안 되고, 정확한 p값과 함께 신뢰 구간을 보고해야 한다.
- 중재 비교에 대한 '결과 요약표'는 문헌고찰한 중재의 상대효과와 절대효과의 크기, 이용 가능한 근거의 양 및 이용 가능한 근거 수준(또는 질)에 관한 주요 정보를 제공한다.
- '결과 요약표'에는 각 중요한 중재결과에 대한 행이 최대 7개까지 포함된다. GRADE의 소프트웨어 GRADEpro GDT를 사용하여 형식화된 '결과 요약표'와 상호적인 '결과 요약표'를 생성할 수 있다.

14.1 자료 분석

이 장에서는 연구결과를 해석하는 핵심 문제를 다룬다. 각 연구결과와 관련된 적용 가능성 및 수치 연구결과 해석에 관한 문제를 보다 자세하게 검토하고 연구결과를 제시하기 위한 방법을 제안한다.

14.1.1 비직접성 및 적용가능성 분석

14.1.1.1 생물학적 변이

생물학적 변이는 다른 인구집단에 대한 연구결과를 적용할 수 있는지에 대한 판단에 영향을 미친다. 생물학적 변이에는 1) 병태 생리의 다양성(예 : 중재에 대한 반응에 영향을 미칠 수 있는 여성과 남성의 생물학적 차이)과 2) 병원체의 다양성(예 : 다양한 기생충에 의해 생기는 말

라리아) 등이 있다. 고찰에서는 이러한 부분에 대해 어디까지 포함시켰는지를 명백히 밝혀야 한다.

14.1.1.2 맥락적 변이

일부 중재, 특히 비약물중재는 일부 맥락에서는 작동하지만 다른 맥락에서는 작동하지 않을 수 있다. 맥락적 요소는 중재를 수행하는 연구자의 전문지식, 경험, 임상의 또는 직원의 관심 우선순위 등과 관련될 수 있다. 좀 더 넓은 맥락 이슈에는 보건 의료 제공자에 대한 수수료나 지불 구조와 같은 시스템 측면을 포함할 수도 있다. 특히 복잡한 중재와 같은 일부 중재는 일부 맥락에서만 부분적으로 구현될 수 있으며, 이러한 상황에서 중재와 구성 요소에 대한 간접적인 판단이 필요하다.

맥락적 요소는 문화적, 언어적 다양성, 사회경제적 지위, 농촌/도시 세팅과 같은 대상 집단 또는 인구집단의 특성과 관련될 수도 있다. 한 세팅에서 다른 세팅으로 일반화할 때 주의해야 한다.

14.1.1.3 준수의 변이

치료 제공자와 수혜자의 준수 변이는 연구결과 적용수준을 제한할 수 있다. 준수의 예측 가능한 차이는 수혜자가 중재를 인식하는 방법의 다양성(예: 유해효과를 중시함), 저소득 국가와 같이 어떤 세팅에 접근할 수 없는 경제 상황 또는 태도 때문일 수 있다. 면밀히 모니터링 되는 무작위배정 임상시험에서 높은 수준가 이루어지더라도 평범한 진료 환경에서 비슷한 수준의 준수로 이어질 것이라고 가정하면 안 된다.

14.1.1.3 가치와 선호도의 변이

보건의료 관리 전략 및 옵션에 관한 결정은 사람들의 가치와 선호도(즉, 결과와 중재에 대한 사람들의 중요성)에 따라 다를 수 있기 때문에 의사 결정이 환자 또는 인구집단의 가치 및 선호도와 일치하는지 확인하는 것이 중요하다. 중재결과에 대한 중요성은 치료받는 사람이 제공되는 옵션을 받아들이는지 또는 받지 않을지에 영향을 미치므로 준수에 영향을 미치는 한 요인이 될 수 있다.

14.1.2 통계 분석의 해석

14.1.2.1 신뢰 구간

개별 연구와 메타분석 결과는 신뢰 구간과 점 추정치가 함께 보고된다. 예를 들어 오즈비가 0.75, 95% 신뢰 구간이 0.70-0.80에서 점 추정치 0.75는 비교 중재와 실험 중재 효과의 크기와 방향에 대한 최선의 추정치이다. 신뢰 구간은 어떤 추정에 내재된 불확실성을 설명하고 진정한 효과가 실제로 존재한다는 것을 합리적으로 확신할 수 있는 범위의 값을 설명한다. 신뢰 구간이 비교적 좁은 경우(예를 들어, 0.70-0.80), 효과 크기는 정확하게 알려져 있는 것이다. 간격이 조금 넓어진 경우에(예 : 0.60-0.93) 불확실 수준은 더 커지지만, 중재의 유용성에

대한 결정을 내리는 데는 여전히 정밀하다. 매우 넓은 간격(예 : 0.50-1.10)은 효과에 대한 지식이 거의 없다는 것을 나타내며, 이 비정밀도는 근거 수준에 영향을 미치고 더 확실한 결론을 도출하려면 추가 정보가 필요하다. 95% 신뢰 구간은 종종 실제 효과가 있음을 95% 확신할 수 있는 범위를 나타내는 것으로 해석된다.

메타분석을 위한 신뢰 구간의 넓이는 표본의 크기에 달려 있다. 적은 수의 연구보다 많은 수의 연구가 더 정확한 효과 크기를 나타낸다. 또한, 변량효과 모형의 경우, 이질성이 증가함에 따라 정밀성이 감소하고 신뢰 구간이 이에 따라 넓어진다. 더 많은 연구가 메타분석에 추가되면 신뢰 구간의 너비는 대개 감소한다. 그러나 추가 연구가 메타분석에서 이질성을 증가시키고 변량효과 모형이 사용되는 경우 신뢰 구간 폭이 증가할 수 있다.

14.1.2.2 p값과 통계적 유의성

p값은 통계 검정의 표준 결과이며 '귀무가설'에 따라 관찰된 효과를 얻을 확률이다. 일반적으로 코크란 리뷰에서 사용되는 두 가지 통계 검정가 있다. 첫 번째는 전반적인 효과(Z-test)이며, 귀무가설은 실험 중재가 비교 중재와 비교하여 전체 관심 결과에 아무런 영향도 미치지 않는다는 것이다. 두 번째는 카이 제곱으로 이질성에 대한 검정으로 귀무가설은 연구들의 중재 효과에 차이가 없다는 것이다.

매우 작은 P 값은 관찰된 효과가 순전히 우연히 일어날 가능성이 거의 없음을 나타내므로 귀무가설 기각이라고 볼 수 있다. 특정 임계값보다 작은지 여부를 검사하여 p값을 해석하는 것이 일반적이다. 특히 0.05 미만의 p값은 종종 "통계적으로 유의한" 것으로 보고되고 귀무가설의 기각을 정당화할 만큼 충분히 작은 것으로 해석된다. 메타분석 결과와 함께 p값을 제시하기로 결정했다면 95% 신뢰 구간과 함께 정확한 p값을 보고해야 한다. 연구결과를 '통계적으로 유의한', '통계적으로 유의하지 않음' 또는 '중요하지 않음'만으로 기술하지 말아야 하며 p값의 임계값에 과도하게 의존해서는 안 되고, 정확한 p값과 함께 신뢰 구간을 보고해야 한다. p값은 일반적으로 두 가지 방법으로 잘못 해석된다. 첫째, 중등도 또는 큰 p값(예를 들어, 0.05 초과)은 중재가 중재결과에 효과가 없다는 근거로 오해될 수 있다. 이 진술과 정확한 해석(결과에 대한 관찰된 효과가 우연으로 인한 것일 확률이 높다) 사이에 중요한 차이가 있다. 이러한 잘못된 해석을 피하기 위해 항상 효과 크기와 95% 신뢰 구간을 제시해야 한다. 두 번째 오해는 요약 효과 크기에 대한 p값이 충분히 작은 경우 실험적 중재가 중요한 이득을 있다고 해석하는 것이다. p값은 실험적 중재 효과가 정확히 있는지 여부에 대한 문제를 나타내지, 그 효과가 중재 대상자에게 중요한 크기인지 여부를 조사하는 것이 아니다.

14.1.2.3 신뢰 구간, 통계적 유의성 및 근거 수준 간의 관계

신뢰 구간 및 비정밀성(imprecision)은 효과 추정에 대한 전반적인 불확실성에 영향을 미치는 영역이다. 비정밀성(즉, 통계적 불확실성)으로 인한 불확실성은 의사결정의 맥락에서 비직접성과 같은 영역만큼 중요 할 수 있다. 따라서 신뢰 구간의 해석은 관심중재 결과에 대한 전반적인 근거의 확실성에 대한 결론에 해당한다. 근거 수준에 영향을 미치는 다른 영역에 대

한 우려가 없는 경우(예: 비뚤림위험, 비직접성) 신뢰구간과 p값의 해석에서 진정한 효과는 전반적 확실성에 대한 것이다. 그러나 근거 수준에 영향을 미치는 다른 영역에 대한 우려가 있는 경우는 진정한 효과에 대한 신뢰 구간 해석은 불확실성의 맥락에서 살펴볼 필요가 있다. 의사결정자는는 평가된 효과 크기(상대, 절대 크기)과 관련하여 이러한 불확실성을 고려해야 한다.

14.1.3 이분형 중재결과 해석

14.1.3.1 상대, 절대 위험 감소

임상의사는 두 가지 중재가 이득(예 : 위험 감소)과 관련될지라도, 사망 위험을 1% 포인트 줄이는 것보다 상대 사망 위험을 25% 줄이는 중재를 처방할 가능성이 더 클 수 있다. 후자는 위험의 상대적인 감소를 의미하고 전자는 위험의 절대적인 감소를 의미한다. 메타분석은 일반적으로 위험비(RR), 오즈비(OR) 또는 위험차(RD)를 사용하여 수행되지만, 결과를 표현하는 몇 가지 다른 방법이 있다.

상대 위험 감소(Relative risk reduction: RRR)는 위험비를 백분율 감소로 재 표현할 수 있는 편리한 방법이다.

$$RRR = 100\% \times (1 - RR)$$

예를 들어 위의 예와 같이 0.75의 위험비는 25%의 상대 위험 감소로 해석된다.

14.1.3.2 필요 치료 수(numbers needed to treat:NNT)

필요 치료 수는 중재의 효과에 대한 정보를 제시하는 일반적인 대안 방법이다. 특정 시간대에 사건을 발생시키거나 피하는데 추가되는 한 명을 위해, 실험 중재를 받아야하는 예상되는 사람의 수로 정의된다. 즉 1명의 치료 효과를 올리기 위해 필요한 치료 환자수를 말한다. 이는 위험차에서 파생되기 때문에, 단일 중재의 일반적인 특성이 아닌 효과의 비교 측정(실험 vs 특정 비교 중재)이다. 또한 필요 치료 수는 '기대값'을 제공한다.

중재가 불리한 결과를 가져오거나 개선보다는 긍정적 결과가 감소할 때, 효과의 방향을 나타내기 위해 '추가적인 위해 중재결과를 위해 필요한 치료수'(number needed to treat for an additional harmful outcome, NNTH)와 같은 문구를 사용하는 것이 좋다.

또한 사건을 참조하므로, 결과가 척도기반 결과의 이분형인 경우, 해당 해석을 신중하게 해야 한다. 예를 들어, 결과가 '없음, 경미, 중등도 혹은 심한' 척도에서 측정된 통증인 경우 '중등도 또는 심각'대 '없음 또는 경미' 한 것으로 이분화 되었을 수 있다. 이러한 자료의 NNT를 '통증 NNT'라고 부르는 것은 부적절하다. 그것은 중등도 또는 심각한 통증을 위한 'NNT'이다.

14.1.3.3 위험차(risk differences)의 해석

'자연적 빈도'를 제시하면(%로 효과를 제시하는 것보다 전체 수로 제시) 통계적 결과에 대한 추론이 더 적절해진다. 예를 들어 '유방암 환자의 치료결과가 1,000명 중 20명에서 1,000명 중 10명으로 감소'로 표현하는 것이 '유방암 위험에서 절대위험 1% 감소'의 표현보다 적절하

다. 특히 사건이 거의 발생하지 않는 경우 확률은 빈도보다 이해하기 어렵다. 절대 위험의 차이를 표현하기 위한 자연적 빈도 표현이 의료 정보 소비자에게 가장 잘 이해된다. 그래서 '결과 요약표'에 중재를 받는 1,000명당 사건이 있는 사람들의 숫자로 절대 위험을 제시한다. 한편 위험차는 기저 사건 비율간 일관성이 거의 없다. 따라서 체계적인 문헌고찰에서 필요 치료수를 계산하는 데는 거의 적합하지 않다.

상대효과 측정(OR 또는 RR)이 메타분석에 선택된 경우, 비교 그룹 위험은 RD 또는 NNT의 계산으로 구체화되어야 한다. 여러 다른 비교 그룹 위험들에서 다른 위험차이(RD), NNT가 발생된 경우에는 가정된 비교 위험(ACR)의 범위에 따라 다양한 RD 또는 NNT로 오즈비와 위험비를 재표현하는 것이 권고된다. 예를 들어, 뇌졸중 비율이 연간 약 12%(1,000 명당 120명)인 고위험 심방 세동 환자 중 와파린은 1,000 명의 환자에게 연간 약 70회 뇌졸중을 예방하는 반면에, 뇌졸중 비율이 연간 약 2%(1000 명당 20명) 인 저위험 심방 세동 환자에서 와파린은 단지 12회 뇌졸중을 예방한다. 이 예를 통해 전형적인 기저 위험이 기대할 수 있는 절대 이익에 미치는 중요한 영향을 이해할 수 있다.

14.1.4 연속형 중재결과 해석

연속형 결과에 대해 문헌고찰 저자는 요약 결과를 제시할 수 있는 여러 가지 옵션을 제공한다. 문헌고찰 저자는 연구 프로토콜에 연속형 결과에 대한 해석 방법을 설명해야 한다.

14.1.4.1 같은 측정을 사용한 연속형 중재결과의 메타분석

이러한 중재결과에 대해 '결과 요약표'에는 두 중재 사이의 평균의 차이가 포함되어야 한다. 그러나 그러한 결과의 단위가 특히 평가 척도(rating scales)와 관련된 경우 해석하기 어려울 수 있다. '결과 요약표'에는 측정 범위의 최소값과 최대값 및 방향이 포함되어야 한다. 최소 중요 차이(MID)를 알면 문헌고찰 저자는 '결과 요약표'의 비고 열을 추가하여 이를 알려야 한다.

14.1.4.2 다양한 측정을 사용한 연속형 결과의 메타분석

다른 도구를 사용한 경우 연속형 자료를 결합하기 위한 메타분석에서 표준화된 평균 차이(SMD)를 사용할 수 있다. 따라서 문헌고찰 저자는 프로토콜 단계에서 분석을 계획할 때 해석 가능성 문제를 고려해야 하며 SMD를 재표현할 적절한 방법이 있는지 또는 평균 비율이나 MID와 같은 대체 효과 측정이 가능한지 고려해야 한다.

14.1.4.3 일반적인 효과 크기 추정치를 사용하여 SMD를 제시하고 해석하기

SMD는 원래 측정 단위가 아닌 표준 단위로 중재효과를 나타낸다. SMD는 중재 집단과 비교 집단 간의 평균 효과의 차이를 합친 표준편차(pooled standard deviation)나 외부 표준편차(연구 결과가 매우 작을 때)로 나눈 값이다. 따라서 SMD의 가치는 효과의 크기(평균의 차이)와 중재결과의 표준 편차에 따라 달라진다. SMD(또는 'Cohen의 효과 크기')를 해석하기 위한 지침 규칙이 존재하며 주로 사회과학 연구에서 제시되었다. 한 예로 0.2는 작은 영향을, 0.5는

중간 정도의 효과를, 그리고 0.8은 큰 효과를 나타낸다. 변형된 예는 〈0.40 는 작음, 0.40 ~ 0.70 중등도, 〉0.70 큰 영향을 의미한다. 그러나 환자의 중요성이 맥락에 따라 다르며 일반적인 진술을 따르지 않기 때문에, 일부 방법론 전문가들은 그러한 해석은 문제의 소지가 있다고 생각한다.

14.1.4.4 익숙한 도구를 사용하여 SMD 재표현하기

SMD를 해석하기 위한 또 다른 방법으로 하나 이상의 특정 측정 도구를 단위로 표현하는 방법이 있다. 이 방법은 가장 친숙한 도구와 관련된 표준편차의 추정치와 SMD를 곱하여 평균의 절대 차이를 계산하는 것이다. 이 표준 편차를 구하기 위해 선택 도구를 사용해 모든 연구의 중재군에 대한 가중 평균을 계산할 수 있다. 사람 간 유사성을 보다 잘 반영하거나 메타분석에서 제시하지 않은 도구를 사용하려면 대표 관찰 연구의 표준 편차를 사용하는 것이 좋다. 따라서 요약 효과는 해당 특정 도구의 원래 단위로 재표현되고 중재 효과의 영향은 친숙한 도구를 사용하여 해석할 수 있다.

14.1.4.5 이분법과 상대 및 절대 측정치로 변환하여 SMD 재표현하기

연속형 측정을 이분법으로 변환하여 이진 척도에 대한 상대적 및 절대적 효과를 계산할 수 있게 한다. 연속 변수가 두 중재군에서 동일한 표준 편차를 갖는 로지스틱 분포를 갖는다는 가정에 기초하여 SMD를(로그)교차비로 변환 할 수 있다, 결과는 근사치로 간주되어야 한다. 로그 오즈 비율(log odds ratio)은 다음과 같이 추정된다.

$$\ln(OR) = \frac{\pi}{\sqrt{3}} SMD$$

(또는 약 1.81 x SMD)

그 결과 오즈비는 '결과 요약표'에서 절대 위험차로 표현되는 추정된 비교 그룹 위험과 결합하여 표시될 수 있다.

14.1.4.6 평균의 비율

관련 연구에서 중재군과 비교 그룹이 모두 같은 방향으로 변하는 경우 변경 점수의 비율을 계산할 수 있으며, 이 비율은 때로 유익할 수 있다. 그러나 중재 비교 그룹이 같은 방향이 아닌 경우처럼 점수 변경에 대한 제한점이 있다. 그리고 비교 그룹 평균이 매우 작기 때문에, 중재군의 차이가 중등도인데도 큰 것으로 중재 비율이 잘못 오도될 가능성이 있다.

14.1.4.7 최소 중요 차이(MID) 단위로 연속형 결과 제시

MID 단위로 결과를 표현하려면 SMD와 같은 방식으로 여러 연구를 결합할 수 있지만, 각 연구의 평균 차이를 표준편차로 나누는 대신 MID로 나눌 수 있다. 표준편차 단위 대신에 합쳐진 결과는 MID 단위를 나타내며 더 쉽게 해석할 수 있다. 이 접근법은 SMD에 의존해서 효과

크기를 왜곡할 수 있는 연구들간 다양한 표준편차 문제를 피할 수 있다. 그러나 이 방법은 잘 구축된 MID를 사용하는 것이 중요하다. 환자의 상당 비율의 환자가 중요한 이득을 얻었을 때, MID보다 작은 차이를 사소한 것으로 해석할 수 있다는 점에서 위험한 접근법이다.

가장 익숙한 도구의 단위로 제시되는 이 접근법은 해당 도구에 대한 광범위한 경험이 있는 경우, 특히 MID가 잘 설정된 경우에 가장 바람직하다.

14.2 결과제시

14.2.1 '결과 요약표'소개

'결과 요약표'는 문헌고찰의 주요 결과를 투명하고 구조적이며 간단한 표 형식으로 제시한다. 특히 근거 수준 또는 근거의 질, 중재 효과 크기, 주요 결과자료의 합과 관련된 핵심 정보를 제공한다. 가장 핵심적인 비교를 나타내는 '결과 요약표'가 적어도 하나는 있어야 한다. 일부 문헌고찰은 하나 이상의 '결과 요약표'를 포함할 수 있다. 예를 들어 문헌고찰이 두 개 이상의 주요 비교를 처리하거나 별도의 표가 필요한 실질적으로 다른 모집단을 포함하는 경우이다. 효과가 다르기 때문에 또는 결과를 별도로 표시하는 것이 중요하기 때문이다.

14.2.2 '결과 요약표'의 중재 결과 선택

'결과 요약표'에 대한 기획은 체계적인 문헌고찰 초기에 문헌고찰 및 '결과 요약표'에 포함될 중재결과를 선택하여 시작한다. 코크란 리뷰는 문헌고찰 질문을 작성하고, 환자 및 다른 의사 결정자에게 중요한 모든 주요 중재결과를 나열한다. 근거 수준을 평가하는 GRADE 접근법은 의사 결정에 핵심적이거나 중요하거나 중요하지 않은 결과로 구분하는 데 도움이 되는 평가 절차이다. 소비자 및 다른 의사 결정권자에게 문헌고찰 프로토콜에 대한 자문과 피드백자을 받으면 이 프로세스를 향상시킬 수 있다.

핵심적 결과(Critical outcomes)에는 중요한 종결점이 포함되어야 한다. 예로 사망 및 주요 이환(예. 뇌졸중 및 심근 경색 등)이 포함된다. 그러나 사소하지만 빈번한 유해효과, 드물지만 중요한 유해효과, 증상, 삶의 질, 치료와 관련된 부담 및 자원 문제(비용)를 나타낼 수도 있다.

의사결정을 위해 모든 환자 중재결과(patient-important outcomes)를 포함하는 질문을 공식화할 때, 이러한 중재결과를 모두 포함하지 않은 연구 보고서에 직면하게 된다. 이것은 특히 유해효과 결과에 해당된다. 예를 들어, 무작위배정 임상시험은 자살 시도와 같은 드문 유해한 결과에 대해서는 보고하지 않을 수 있다. 모든 중요한 중재결과에 대한 자료를 얻으려면 비임의 연구를 조사해야 할 수도 있다.

무작위배정 임상시험에서 연구되지 않은 희귀하거나 장기간의 유해효과를 조사하기 위해 비 임의 연구를 포함하는 것이 좋다. 이것은 위해 결과가 이득 분석에서 사용된 연구대상자와 다른 연구대상자일 가능성이 높다. 문헌고찰 저자는 그런 차이가 연구결과에 얼마나 영향을

미칠지 고려해야 한다. 이것은 연구대상자와 관련된 간접적인 관계로 인해 근거 수준에 영향을 줄 것이다.

비임의 연구는 무작위배정 임상시험이 중재결과에 대해 보고하지 않을 때와 무작위배정 임상시험에서 나온 근거가 매우 낮게 평가될 때 중요한 정보를 제공해 줄 수 있다.

14.2.3 '결과 요약표'의 일반 템플릿

의사 결정자가 필요로 하는 가장 핵심적인 정보를 포함하고 최적의 제시를 제공하고 문헌고찰 전반에 걸쳐 일관성과 사용 용이성을 보장하기 위해 '결과 요약표'의 몇 가지 대안으로 표준 버전이 개발되었다(그림 14-1, 14-2). GRADE 방식을 지원하기 위해 개발된 GRADE의 공식 소프트웨어 패키지(GRADEpro GDT, www.gradepro.org)가 이용가능하다.

코크란에서 '결과 요약표'에는 다음과 같은 요소가 있다.

- 이용 가능한 근거의 연구대상자와 세팅에 대한 간단한 설명(문헌고찰 질문에서 정의한 것과 약간 다를 수도 있고 좁을 수도 있음)
- 실험 및 비교 중재를 포함하여 '결과 요약표'에서 나타낸 비교에 대한 간략한 설명
- 핵심적인 혹은 중요한 7개까지의 중재결과 목록.
- 각 중재결과에 대한 전형적인 부담의 측정(예 : 비교 중재에 대한 예시적 위험(illustrative risk) 또는 예시적 평균(illustrative mean))
- 측정된 효과의 절대 및 상대적 크기(둘 다 가능한 경우).
- 연구대상자 수와 각 중재결과의 분석에 기여한 연구 수
- 각 중재결과 에 대한 전체적 근거 수준에 대한 GRADE 평가(결과에 따라 다를 수 있음)
- 코멘트 공간.
- 설명(이전에는 각주라고 함)

이상적으로 '결과 요약표'는 보다 자세한 표인 '근거 프로파일'에 의해 만들어지고 이 둘은 서로 링크된다. 근거 프로파일은 똑같은 중요한 건강 중재결과를 나열하고 근거 수준과 연구결과에 대한 개별 고려 사항에 대해 '결과 요약표'보다 더 자세하게 설명한다. 근거 수준을 평가하기 위해 구조화된 접근 방식을 사용한다. 문헌고찰 저자는 GRADEpro GDT의 근거 프로파일에서 근거 수준 평가를 완료함으로써 '결과 요약표'를 보다 쉽게 개발할 수 있다. 그런 다음 자동으로 이것을 GRADEpro GDT에서 '결과 요약' 형식 중 하나로 변환할 수 있다.

이분형 중재결과에 대한 효과의 크기를 측정하기 위해 '결과 요약표'는 상대 위험도 및 절대 위험 측정치을 제공해야 한다. 다른 유형의 자료의 경우, 절대값 측정(예 : 연속형 자료의 평균값의 차이)만으로 충분할 수 있다. 효과의 크기가 의미 있는 방식으로 제시되는 것이 중요하며, 메타분석 결과의 일부 변형이 필요할 수 있다. 주요 비교가 둘 이상인 문헌고찰에는 각 비교에 대한 '결과 요약표'가 별도로 포함되어야 한다.

그림 14-1은 '결과 요약표'의 예이다. 그림 14-2는 문헌고찰 결과에 대한 사용자의 이해와 해석을 용이하게 하는 대체 형식을 제공한다

장시간 비행기 탑승객을 대상으로 한 압박 스타킹 착용과 미착용 비교
대상자: 장시간 비행자(적어도 6시간 이상)
세팅: 국제 항공 여행 **중재 :** 압박 스타킹[a] **비교:** 압박스타킹 착용 안함

Outcomes	Illustrative comparative risks* (95% CI)		Relative effect (95% CI)	Number of participants (studies)	Certainty of	Comments
	Assumed risk	Corresponding risk				
	Without stockings	**With stockings**				
Symptomatic deep vein thrombosis (DVT)	See comment	See comment	Not estimable	2821 (9 studies)	See comment	0 participants developed symptomatic DVT in these studies
Symptomless DVT	**Low risk population**[b]		**RR 0.10** (0.04 to 0.26)	2637 (9 studies)	⊕⊕⊕⊕ **High**	
	10 per 1000	**1 per 1000** (0 to 3)				
	High risk population[b]					
	20 per 1000	**2 per 1000** (1 to 8)				
Superficial vein thrombosis	**13 per 1000**	**6 per 1000** (2 to 15)	**RR 0.45** (0.18 to 1.13)	1804 (8 studies)	⊕⊕⊕◯ **Moderate**[c]	
Oedema Post-flight values measured on a scale from 0, no oedema, to 10, maximum oedema	The mean oedema score ranged across control groups from **6 to 9**	The mean oedema score in the intervention groups was on average **4.7 lower** (95% CI –4.9 to –4.5)		1246 (6 studies)	⊕⊕◯◯ **Low**[d]	
Pulmonary embolus	See comment	See comment	Not estimable	2821 (9 studies)	See comment	0 participants developed pulmonary embolus in these studies[e]
Death	See comment	See comment	Not estimable	2821 (9 studies)	See comment	0 participants died in these studies
Adverse effects	See comment	See comment	Not estimable	1182 (4 studies)	See comment	The tolerability of the stockings was described as very good with no complaints of side effects in 4 studies[f]

*The basis for the **assumed risk** is provided in footnotes. The **corresponding risk** (and its 95% confidence interval) is based on the assumed risk in the intervention group and the **relative effect** of the intervention (and its 95% CI).
CI: confidence interval; RR: risk ratio; GRADE: GRADE Working Group grades of evidence (see explanations).

a 이 문헌고찰에 포함된 9개의 연구에서 스타킹은 무릎아래 압축 스타킹이었다. 4개 연구에서 압축 강도는 발목에서 20 mmHg-30 mmHg였다. 다른 4개의 연구에서 10mmHg- 20 mmHg였다. 스타킹은 다양한 크기로 제공된다. 스타킹이 무릎 주위에 너무 꽉 끼 이는 경우 필수적인 정맥으로의 복귀를 방지하여 혈액이 무릎 주위에 쌓이게 할 수 있다. 압축 스타킹은 올바르게 장착되어야 한다. 너무 단단한 스타킹은 장거리 비행시 피부에 손상을 줄 수 있으며 잠재적으로 궤양을 유발하고 DVT의 위험을 증가시킨다. 편안한 착용감을 위해 여행하기 전에 집에서 스타킹을 착용하는 것이 좋다. 연구대상자들은 대부분의 연구에서 비행기 출발 2-3시간 전에 스타킹을 착용했다. 스타킹의 가용성 및 비용은 다를 수 있다.
b 2개 연구에서 DVT 이전 경험, 응고 장애, 심한 비만, 뼈 또는 관절 문제로 인한 이동성의 제한, 지난 2년 이내의 신생 질환, 큰 정맥류의 고위험 연구대상자를 모집했으며, 연구 중 한개에서 연구대상자들은 190cm보다 키가 크고 몸무게 90kg이상 이었다. 위험도가 높은 연구대상자를 제외한 7개 연구에서 발생률은 1.45% 였고 위험도가 높은 연구대상자(최소한 하나의 위험 요소 포함)를 모집 한 두 연구에서 발생률은 2.43%였다. 각각 다른 위험 층(strata)을 표현하기 위해 1000당 10과 30을 사용했다.
c 신뢰 구간은 차이가 없으며 작은 증가(small increase)를 배제하지는 않는다.
d 부종의 측정은 검증되지 않았거나(결과의 비직접성) 중재(비뚤림위험)에 대해 눈가림하지 하지 않았다.
e 사건발생이 거의 없고 연구대상자의 수가 많은 경우, 근거의 질(특히 비정밀성에 대한 판단)에 대한 판단은 절대적 효과에 근거 할 수 있다. 결과가 적절하게 평가되고 실제로 사건이 2821명의 연구대상자에서 발생하지 않았다면, 질 등급은 '높음'이 될 수 있다.
f 한 연구에서 무릎 부위 정맥류 혈전증 4개 사례를 제외하고는 유해효과가 보고된 다른 연구가 없었다.

그림 14-1. '결과 요약표' 예

소아 항생제에 대한 보조치료제로 프로바이오틱스 투여와 비투여 비교
대상자: 항생제 투여 어린이
세팅: 입원 외래 환자, 중재: 프로바이오틱스, 비교: 프로바이오틱스 없음

Outcomes No of participants (studies)	Relative effects (95% CI)	Anticipated absolute effects* (95% CI) Without probiotics	 With probiotics	 Difference	Certainty of the evidence (GRADE)	Comments
Incidence of diarrhoea: Probiotic dose 5 billion CFU/day Follow-up: 10 days to 3 months					⊕⊕⊕⊖ **moderate**[b] Due to risk of bias	Probably decreases the incidence of diarrhoea.
Children < 5 years 1474 (7 studies)	**RR 0.41** (0.29 to 0.55)	**Children < 5 years** **22.3%**[a]	**8.9%** (6.5 to 12.2)	**13.4% fewer children**[a] (10.1 to 15.8 fewer)		
Children > 5 years 624 (4 studies)	**RR 0.81** (0.53 to 1.21)	**Children > 5 years** **11.2%**[a]	**9%** (5.9 to 13.6)	**2.2% fewer children**[a] (5.3 fewer to 2.4 more)	⊕⊕⊖⊖ **low**[b, c] Due to risk of bias and imprecision	May decrease the incidence of diarrhoea.
Adverse events[d] Follow-up: 10 to 44 days 1575 (11 studies)	-	**1.8%**[a]	**2.3%** (0.8 to 3.8)	**0.5% more adverse events**[e] (1 fewer to 2 more)	⊕⊕⊖⊖ **low**[f, g] Due to risk of bias and inconsistency	There may be little or no difference in adverse events.
Duration of diarrhoea Follow-up: 10 days to 3 months 897 (5 studies)	-	The mean duration of diarrhoea without probiotics was **4 days.**	-	**0.6 fewer days** (1.18 to 0.02 fewer days)	⊕⊕⊖⊖ **low**[h, i] Due to imprecision and inconsistency	May decrease the duration of diarrhoea.
Stools per day Follow-up: 10 days to 3 months 425 (4 studies)	-	The mean stools per day without probiotics was **2.5 stools per day.**	-	**0.3 fewer stools per day** (0.6 to 0 fewer)	⊕⊕⊖⊖ **low**[j, k] Due to imprecision and inconsistency	There may be little or no difference in stools per day.

*The basis for the **risk in the control group** (e.g. the median control group risk across studies) is provided in footnotes. The **risk in the intervention group** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI). **CI:** confidence interval; **RR:** risk ratio.

설명
a 비교 집단의 위험 추정치는 비교군의 통합(poolded) 추정치로부터 나온다. 상대효과는이용 가능한 사례 분석을 기반으로 한다.
b 높은 추적자료 손실로 인한 비뚤림위험이 높다.
c 사건이 적고 신뢰 구간이 상당한 이득이나 위해를 포함하여 비정밀성 발생.
d 유해효과 : 발진, 오심, 고장(flatulence), 구토, 객담 증가, 흉통, 변비, 미각 장애, 식욕 저하
e 위험은 합쳐진(pooled) 위험차로부터 계산되었다.
f 높은 비뚤림위험. 16 건 중 11 건만이 유해효과에 대해 보고되어 선택적 보고 비뚤림
g 심각한 비일관성. 상대적으로 적은 수의 임상시험에서 많은 프로바이오틱스제와 용량이 평가되어, 결론에서 많은 프로바이오틱스제와 용량의 안전성에 대한 결론을 이끌어 내는데 제한적이다.
h 심각한 설명되지 않은 비일관성(큰 이질성 I^2 = 79 %, P 값 [P = 0.04], 점 추정(point estimates)과 신뢰 구간은 상당히 다양함)
i 심각한 비정밀성. 설사가 0.02일 보다 적은 상한선은 환자에게 중요한 것으로 간주되지 않는다.
j 심각한 설명되지 않은 비일관성(큰 이질성 I^2 = 78 %, P 값 [P = 0.05], 점 추정(point estimates)과 신뢰 구간은 상당히 다양함)
k 심각한 비정밀. 95 % 신뢰 구간에서 효과가 없고, 하루 0.60회 설사보다 적은 하한이 환자에게 중요한지 의문스럽다.

그림 14-2. 대안적인 '결과 요약표'예

14.2.4 '결과 요약표' 생성하기

GRADE의 소프트웨어인 GRADEpro GDT(www.gradepro.org)는 '결과 요약표' 준비 과정에서 작성자를 도울 수 있다. GRADEpro는 실험 중재와 관련된 상대효과와 절대 위험을 산출하기 위해 비교 그룹 위험 및 효과 추정(문헌고찰 저자가 입력하거나 RevMan에서 생성된 파일에서 가져온 데이터)을 사용할 수 있다. 또한 사용자에게 GRADE 평가 프로세스를 유도하고, 문헌고찰시 독립형 표로 사용할 수 있는 표(RevMan과 같은 소프트웨어로 직접 가져 오기를 포함) 또는 '결과 요약' 표(GRADEpro의 도움말 자료 참조)를 만들 수 있게 한다.

14.2.5 '결과 요약표'의 통계적 고려 사항

14.2.5.1 이분형 결과

'결과 요약표'에는 이분형 결과에 대한 효과의 절대 및 상대 측정치 모두 포함되어야 한다. 위험비, 오즈비 및 위험차는 두 그룹의 이분형 중재결과 자료를 비교하는 다양한 방법이다. 또한 어떤 사건(예 : '예' 또는 '아니요')이 분석의 초점에 따라 두 가지의 위험 비율이 있다. 메타분석은 일반적으로 효과의 측정치로 위험비 또는 오즈비를 사용해야 한다. 따라서 상대 효과의 단일 측정치는 절대 효과의 단일 측정치보다 더 적절한 요약일 수 있다. 상대 효과가 실제로 여러 연구에서 일관성이 있다면, 다른 비교 그룹의 위험은 다른 영향을 가질 것이다.

'결과 요약표'는 일관된 상대 효과를 전제로 한다. 따라서 다른 비교 그룹 위험에 대한 이 효과의 영향을 고려하는 것이 중요하며 이는 예후적 근거에 대한 근거 수준 평가가 필요할 수 있다. 비교 그룹 위험에서 메타분석 위험비 또는 오즈비로부터 해당 중재군 위험을 추정하는 것이 가능하다. '해당 위험'열에 제공된 숫자는 인접 열의 '위험'에만 해당된다.

메타분석 위험비(RR)와 추정 비교군 위험도(ACR: assumed comparator risk)에 대해 해당 중재 위험은 다음과 같이 얻어진다.

$$1000\text{명당 해당 중재의 위험} = 1000 \times \frac{OR \times ACR}{1 - ACR + (OR \times ACR)}$$

14.2.5.2 사건발생 시간 중재결과

비교 집단에 대한 실험적 중재가 시간 대 사건 결과에 미치는 영향은 일반적으로 위험비(HR)를 사용하여 측정된다. 위험비는 상대 효과 크기를 나타낸다. 특정 집단에 대해 절대 위험 및 기타 해석 가능한 양을 얻기 위해 다양한 방법으로 사용될 수 있다. 모든 방법은 일관된 상대 효과(즉, 위험 비율은 시간에 따라 변하지 않음)을 가정하여 작성된다.

(i) 특정 기간 내 무사건 생존의 절대 위험 : 무사건 생존으로 측정된 시간 대 사건 결과에 대한 절대 효과를 얻기 위해, 비교 집단에서 사건이 없는 환자의 추정 비율(R)과 함께 요약 위험비(HR)를 사용할 수 있다.

$$1000\text{명당 해당 중재의 위험} = 1000 \times \exp(\ln R \times HR)$$

(ii) 특정 기간 내에 일어날 수 있는 사건의 절대 위험: 이 절대적 효과를 얻으려면 요약 HR을 다시 사용할 수 있다.

$$1000\text{명당 해당 중재의 위험} = 1000 - 1000 \times \exp(\ln(1-R) \times HR)$$

(iii) 사건의 중간 시간(Median time to the event): 절대 숫자 대신에, 중재와 비교 그룹에서 사건의 시간은 몇 달 또는 몇 년의 중간 생존 시간(median survival time)으로 표현될 수 있다. 중간 생존 시간을 얻기 위해 비교 집단에서 가정된 생존 기간에 합쳐진 위험비를 적용할 수 있다.

해당 중간 생존(개월) = 비교 그룹 중간 생존 시간(개월) / 위험비(HR)

시간 대 사건 분석의 결과를 재표현하기 위한 이 세 가지 옵션 모두에 대해, 상응하는 중재 위험에 대한 상한 하한 신뢰 한계는 위험비(HR)를 상한 신뢰 한계와 하한 신뢰 한계로 대체함으로써 얻어진다.

14.2.6 '결과 요약표'의 상세 내용

14.2.6.1 표 제목 및 헤더

각 '결과 요약표'의 제목은 연구대상자 측면에서 보건의료 질문을 구체화하고 중재의 비교가 정확히 무엇인지 명확히 해야 한다.

각 '결과 요약표'의 첫 행은 다음 정보를 제공해야 한다.

- 환자 또는 연구대상자 : 이것은 관심 연구대상자(및 하위군)와 중재가 유도하는 가장 핵심적인 부정적인 결과의 위험 크기를 이상적으로 명확하게 한다.
- 세팅 : 결과 요약의 적용 가능성을 다른 세팅으로 제한할 수 있는 보건의료 질문에 대한 세팅의 특성을 구체적으로 나타내야 한다.
- 중재 : 실험적인 중재.
- 비교 : 비교 중재(특정 중재 없음을 포함)

14.2.6.2 중재결과

'결과 요약표'의 행에는 의사결정에 필수적인 모든 바람직하고 바람직하지 않은 건강 결과

(중요도 순으로 나열)가 포함되어야 한다(최대 7가지 결과). 문헌고찰 저자는 결과 측정의 시간 프레임(예 : 90일 또는 12개월)과 도구 점수 유형(예 : 0에서 100까지)을 제공해야 한 다.

문헌고찰 저자는 자료의 사용 가능 여부에 관계없이 미리 지정된 중대하고 중요한 결과를 표에 포함시켜야 한다. 그러나 프로토콜의 작성이나 분석이 수행된 후에만 결과의 중요성(예 : 심각한 악영향)이 알려질 수 있을 가능성에 주의를 기울여야 하며, '결과 요약표'에 이를 포함시키는 적절한 조치를 취해야 한다.

'결과 요약표'는 다른 비교 위험과 효과 크기에 대해 집단의 하위군에서의 영향을 개별적으로 포함할 수 있다. 예를 들어, 그림 14-2에서 5세 미만의 아동과 별도로 5세 이상의 아동에 대한 효과가 제시된다. 문헌고찰 저자는 다른 집단에 대한 '결과 요약표'를 별도로 생성할 수도 있다.

심각한 유해효과를 포함해야 하지만, 작은 유해효과들을 단일 결과로 결합하여 각주에 설명하는 것이 가능할 수 있다.

일반적으로 표를 단순하게 유지하기 위해 의사 결정에 중요한 결과에 대해서만 여러 시점을 제시해야 한다. 나머지는 가능한 공통 시점에서 제시되어야 한다.

'결과 요약표'에 연속형 결과 측정을 제시할 수 있으며 단위가 명확하고 쉽게 해석할 수 있어야 한다. 그러나 많은 측정 도구가 쉽게 해석되지 않는 경우가 많다. 이를 위해 50% 이상 개선과 같이 이분형 결과로 변환하여 제시하는 것이 더 바람직할 수 있다.

14.2.6.3 비교 중재 위험 측정

비교 중재를 받는 연구대상자에게 측정하는 경우 여러 가지 형태로 위험을 제시한다. 이분형 중재결과의 경우 중재결과의 빈도에 따라 100명 또는 1000명 당 사건을 경험하는 사람들의 수로 표시하는 것이 좋다. 연속형 중재 결과의 경우, 측정된 중재결과의 평균값 또는 중간값으로 명시된다.

추정된 비교 중재 위험도는 다양한 환자 집단의 전형적 위험 평가를 기반으로 하는데 이들 자료는 문헌고찰 자체, 문헌고찰에서 선정된 연구 중 대표적인 것 예후 연구 또는 다른 근거자료의 체계적인 문헌고찰에서 얻을 수 있다.

설명문 각주는 각 비교 그룹 위험도에 대한 출처 또는 근거를 기술해야 한다. 그림 14-1에서 임상의는 심부 정맥 혈전증의 위험 요인을 가진 사람들과 그렇지 않은 사람들을 쉽게 구분할 수 있다. 기저상태 위험에 변이가 거의 없는 것으로 알려진 경우 문헌고찰 저자는 여러 연구에서 중간 비교 그룹 위험도(median comparator group risk)을 사용할 수 있다. 일반적인 위험 요소를 알 수 없는 경우, 고위험연구대상자는 선정된 연구에서 제공된 두 번째로 높은 위험도에서 저위험연구대상자는 두 번째로 낮은 위험연구대상자로부터 위험을 선택한다.

14.2.6.4 중재 위험도(Risk with intervention)

이분형 결과의 경우, 문헌고찰 저자는 신뢰 구간과 각 비교 그룹 위험에 대해 상응하는 절

대 위험을 제공해야 한다. 이 절대 위험은 보통 상대 효과 열에 제시된 메타분석 결과로부터 도출된다. 예를 들어 1,000명당 사건을 경험하는 사람 수와 같이, 비교 중재 위험과 동일한 형식으로 절대 효과를 나타내야 한다.

연속형 결과의 경우 평균의 차이 또는 평균의 표준화된 차이를 신뢰 구간과 함께 제시해야 한다. 이는 전형적으로 메타분석으로부터 직접 얻어진다. 그림 14-.1, 그림 14-2와 같이 의미를 명확히 하기 위해 설명 텍스트를 사용해야 한다.

14.2.6.5 위험차

이분형 결과의 경우, 추가 선택 사항으로 '결과 요약표' 형식 중 하나를 사용하여 위험차를 제공 할 수 있다(그림 14-2 참조). 이 위험차는 실험과 비교 중재 사이의 차이를 나타내며 일반적으로 상대 효과 열에 제시된 메타분석 결과에서 파생된다. 비교 중재의 위험에 상응하는 같은 형식으로 위험 차이를 제시해야 한다. 연속 결과의 경우 '결과 요약표'에 이 옵션이 포함된 경우 평균 차이가 여기에 표시되고 '해당 위험' 열은 공백으로 남긴다(그림 14-2 '결과 요약표' 참조).

14.2.6.6 상대 효과(95% CI)

상대 효과는 일반적으로 동일한 효과 측정을 기반으로 수행된 메타분석에서 얻어진 95% 신뢰 구간과 위험비(RR) 또는 오즈비(또는 hazard ratio)가 될 것이다. 위험비와 오즈비는 비교 중재 위험이 낮고 효과가 적을 때 비슷하지만, 비교중재 위험도가 증가할 때는 상당히 다를 수 있다. 메타분석은 문헌고찰 저자가 판단하여 고정 효과 또는 변량효과의 가정 중 하나를 포함할 수 있고, 상대 효과는 중재 효과 크기 또는 연구간 중재의 평균 효과 크기 중 하나를 포함한다.

14.2.6.7 연구대상자(연구) 수

이 열에는 각 결과에 대해 선정된 연구에서 평가된 연구대상자 수와 해당 연구 수가 포함되어야 한다.

14.2.6.8 근거 수준(GRADE)

문헌고찰 저자는 근거 수준에 대해 논평해야 한다. 문헌고찰 저자는 GRADE그룹이 개발된 특정 근거 수준 시스템을 사용해야 한다. GRADE 접근법은 결과에 따라 '높음', '보통', '낮음' 또는 '매우 낮음'으로 전체적 근거 수준을 분류한다. 이것은 판단의 결과이지만, 판단 프로세스는 투명하게 이루어져야 한다.

예를 들어, 비뚤림위험이 낮은 몇 개의 무작위배정 임상시험의 요약이 있다면 근거 수준은 '높음'이지만 비뚤림위험, 비일관성, 간접성(비직접성), 비정밀또는 출판 비뚤림 우려가 있는 경우 근거의 질 등급이 낮아진다. '높음'이외의 판단은 설명의 각주 또는 '결과 요약표'의 '코멘트'열을 사용하여 설명하도록 한다.

14.2.6.9 코멘트

'코멘트' 영역의 목적은 행에 식별된 정보 또는 데이터를 해석하는 데 도움을 주는 것이다. 예를 들어, 이것은 결과 측정의 타당성이나 효과의 크기와 관련된 변수에 대한 것일 수 있다. 결과에 대한 중요한 주의 사항을 여기에 표시해야 한다. 모든 행에 코멘트가 필요하지는 않으며 코멘트가 필요한 부분이 없으면 공백으로 두는 것이 가장 좋다.

14.2.6.10 설명

전반적인 GRADE 평가와 같은 '결과 요약표'의 판단을 뒷받침하는 자세한 설명이 각주로 포함되어야 한다. 설명은 내용의 중요한 측면에 대한 이론적 근거를 설명해야 한다. 설명은 간결하고 유익하며 이해하기 쉽고 정확해야 한다. 설명이 각주에 충분하게 설명될 수 없는 경우 문헌고찰 저자는 문헌고찰 결과 및 논의에서 자세한 내용을 제공해야 한다.

14.3 자료분석과 결과제시 관련 MECIR

C72: 연구결과 해석(필수)	
연구결과를 통계적으로 '유의한' 또는 '유의하지 않은'으로 기술하지 않는다. 신뢰 구간과 폭으로 해석한다.	저자는 흔히 '효과의 근거가 부족하다'는 사실을 '효과가 없다'로 근거를 잘못 해석한다.

14.4 참고문헌과 읽을거리

- Schünemann HJ, Higgins JPT, Vist GE, Glasziou P, Akl EA, Skoetz N, Guyatt GH. Chapter 14: Completing 'Summary of findings' tables and grading the certainty of the evidence. Draft version(29 January 2019) for inclusion in: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA(editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.0(updated July 2019). Cochrane, 2019. Available from www.training.cochrane.org/handbook.
- Schünemann HJ, Vist GE, Higgins JPT, Santesso N, Deeks JJ, Glasziou P, Akl EA, Guyatt GH. Chapter 15: Interpreting results and drawing conclusi ons. Draft version(29 January 2019) for inclusion in: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA(editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.0(updated July 2019). Cochrane, 2019. Available from www.training.cochrane.org/handbook.
- Schüunemann HJ, Higgins JPT, Vist GE, Glasziou P, Akl EA, Skoetz N, Guyatt GH. Chapter 14: Completing 'Summary of findings' tables and grading the certainty of the evidence. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions

version 6.0 (updated July 2019). Cochrane, 2019. Available from www.training.cochrane.org/handbook.

제15장 근거 수준 평가와 결론의 도출

핵심 포인트

- 코크란 그룹에서도 전체적 근거 수준(또는 질)을 평가하기 위해 권고 평가, 개발 및 등급화를 위한 GRADE 접근법을 채택하였다.
- GRADE 접근법은 주어진 중재결과에 대한 전체적 근거 수준에 대해 네 가지 수준, 즉 높음, 중등도, 낮음 및 매우 낮음으로 평가한다.
- 비뚤림위험, 비일관성, 간접성, 비정밀 및 출판 비뚤림의 5가지 영역을 고려하여 GRADE의 근거 수준에 대한 평가가 결정된다. 비무작위 연구와 준 무작위 연구(rarely randomized studies)의 근거에 대해서는 세 가지 추가 영역을 고려하여 등급을 상향시킬 수 있다.

15.1 근거 수준 평가

15.1.1 GRADE 접근법

GRADE 그룹은 근거 수준을 평가하는 시스템을 개발했다. 세계 보건기구(WHO), 미국 의사 회, 미국 혈액 학회, 캐나다 의약청 및 건강 기술국(CADTH) 및 미국의 NIH, 영국의 NICE를 포함한 100개 이상의 조직에서 GRADE 시스템(www.gradeworkinggroup.org)을 채택했다.

체계적 문헌고찰에서 활용하는 GRADE 접근법은 근거 수준을 "효과 추정치가 실제 크기와 동일하다고 확신 하는 정도"로 정의한다. 근거 수준을 평가하는 것은 비뚤림위험(방법론적 질), 비일관성, 근거의 간접성, 효과 추정의 비정밀도 및 출판 비뚤림위험을 고려해야 한다. GRADE 시스템은 각 중재결과에 대한 전체적 근거 수준도 평가한다. 근거 수준을 결정하는 영역에 대한 판단 내용은 결과나 고찰 파트 혹은 '결과 요약표'에 일부 내용을 기술해야 한다.

GRADE 접근법은 4가지 수준을 규정한다(그림 15-1). 진단 검사를 포함한 중재의 경우 근거 수준을 평가하기 위한 출발점은 포함된 연구설계를 무작위배정 임상시험과 비무작위 중재연구(NRSI)의 두 가지 유형으로 구분하는 것이다. 잠재적 위해를 평가할 때 비무작위 연구의 정보에 의존할 수 있으며, 무작위배정 임상시험과 비무작위 연구 근거가 상호 보완적일 수 있다.

GRADE에서는 무작위배정 임상시험의 근거를 높은 수준으로 평가하는 반면 비무작위 연

구의 근거는 낮은 근거 수준으로 시작한다. 비무작위 연구가 낮은 등급을 받는 이유는 무작위 배정의 부재로 인한 잠재적 비뚤림(즉, 교란과 선택 비뚤림) 때문이다.

그러나 ROBINS-I 도구를 사용하면, 모든 연구 유형에서 근거 수준이 높은 것으로 시작할 수 있다. 모든 연구 디자인을 높은 수준으로 시작하는 접근법은 비무작위 연구를 낮은 근거 수준으로 평가하는 GRADE 접근법과 충돌하지 않는다. 비무작위 연구의 근거 수준은 일반적으로 무작위배정을 하지 않아 교란과 선택 비뚤림과 관련된 내재된 비뚤림위험 때문에 2단계 수준이 하향되어야 하기 때문이다. 비무작위 연구를 높은 확신에서 낮은 확신으로 낮추지 않으려면 교란과 선택 비뚤림에 대한 우려를 완화시키는 것에 대한 투명하고 상세한 근거제시가 필요하다.

가장 높은 수준의 등급은 그림 15-1에 열거된 모든 GRADE 요인에 아무런 우려가 없을 때의 근거이다. 문헌고찰 저자는 그림 15-1의 다섯 가지 요인에 따라 근거를 중등도, 낮음 또는 매우 낮음 수준의 근거로 하향 평가하는 경우가 많다. 일반적으로 근거 수준의 등급은 각 요인에 대해 한 단계, 모든 요인에 대해 최대 세 단계까지 떨어진다. 어느 한 영역에서 매우 심각한 문제가 있는 경우(예 : 비뚤림위험을 평가할 때 모든 연구에서 배정은폐가 되지 않음, 눈가림이 이루어지지 않음, 환자의 50% 이상이 탈락함) 2단계 하향 평가될 수 있다.

문헌고찰 저자는 일반적으로 ROBINS-I을 사용하더라도 낮은 수준으로 비무작위 연구결과를 평가할 것이다. 그러나 그러한 비무작위 연구에서 효과의 크기가 크며, 그 효과에 대한 명백한 비뚤림이 없으면 문헌고찰 저자는 근거 수준으로 중등도, 효과의 크기가 충분히 크다면, 심지어 높은 수준으로 평가할 수 있다(그림 15-1). 매우 낮은 수준은 연구에 중요한 문제가 있고 체계적이지 않은 임상 관찰(예 : 사례 시리즈 또는 사례 보고서)이 있는 연구에 해당한다.

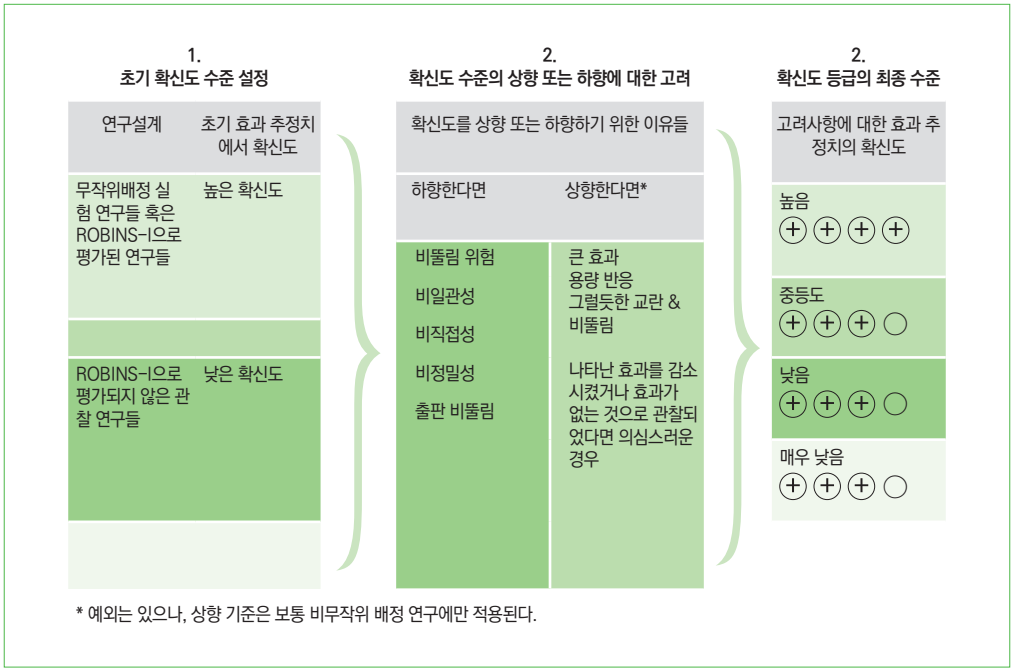


그림 15-1. GRADE 접근의 근거 수준

15.1.2 전체적 근거 수준을 하향조정하는 영역

특정 결과에 대한 근거 자료의 수준을 낮추기 위한 다섯 가지 영역을 보다 자세하게 설명한다. 각각의 경우에, 근거를 하향 평가하는 이유가 발견되면, '그 강도를 '제한 없음'(하향 평가할 만큼 중요하지 않음), '높음'(근거 수준을 한 단계 낮춤) 또는 '매우 높음(근거 수준을 2단계 낮춤)'으로 분류한다. ROBINS-I으로 평가한 비무작위 연구의 경우 '매우 심각함'으로 분류되면 3단계를 낮추어야 한다.

1) 비뚤림위험이나 제한점 : 연구결과에서 중재 효과의 비뚤림위험 평가에서 주요 제한점이 있다면, 효과 추정치에 대한 확신도는 감소한다. 무작위배정 임상시험에서 이러한 방법론 적 제한점으로는 무작위 순서 생성 실패, 배정 순서 은폐의 부재, 눈가림 부재(특히 주관적 중재결과), 다수의 추적 결과 누락 또는 선택적 연구결과 보고 등이 있다.

무작위배정 임상시험의 비뚤림위험은 '낮음', '일부 우려', '높음'으로 제시하고, ROBINS-I으로 평가한 비무작위 연구의 비뚤림위험 평가는 '낮음'. '중등도', '높음' '매우 높음' 수준으로 분류된다. 특히 비뚤림위험이 '낮음'은 '제한점 없음'에 해당되고, ,'일부 우려'는 '제한 없음' 또는 '높음'을, '높음'은 '높음' 또는 '매우 높음'을 나타낸다. ROBINS-I에서 비뚤림위험이 '매우 높음'은 매우 심각한 제한이 있음을 나타낸다. 문헌고찰 저자는 잠재적인 비뚤림 정도에

따라 적절한 판단을 하여 카테고리를 결해야 한다.

모든 연구는 어느 정도 비뚤림위험이 있다. 문헌고찰 저자는 연구 제한점에 대한 비뚤림으로 근거 수준을 하향등급 해야 할지에 대한 전반적인 판단을 내려야 한다. 연구 제한점의 평가는 모든 연구보다는 '결과 요약표'의 중재결과에 기여하는 연구에 적용해야 한다.

그림 15-1은 '결과 요약표'에 포함된 각 결과에 대한 판단은 연구 제한점에 대한 판단을 위한 비뚤림위험 평가에서부터 이루어져야 한다. 높은 근거 수준 등급은 대부분의 근거가 비뚤림위험이 낮은 기준에 부합하는 연구에서 나온 경우에만 달성될 수 있다.

근거 수준은 대부분의 근거가 하나의 항목에 대한 결정적인 제한점 또는 여러 항목에 대한 제한점이 있는 개별 연구들에서 오는 경우 한 단계씩 낮아질 수 있다.

2) 설명되지 않는 이질성 또는 결과의 비일관성: 연구결과의 추정치가 연구마다 크게 다를 경우(결과의 이질성 또는 변동성) 연구자는 비일관성의 이유를 찾아야 한다. 예를 들어, 의약품은 더 아픈 인구집단이나 더 많은 용량을 투여할 때 상대 효과가 더 클 수 있다. 중요한 효과변경요인이 존재하고 중요한 중재결과가 서로 다른 하위 그룹에서 다르다는 근거가 있는 경우, 별도의 집단에 대해 '결과 요약표'를 별도로 고려할 수 있다. 비일관성이 존재하고 결과 해석에 영향을 미치지만 이용 가능한 자료로 가능한 설명을 확인할 수 없는 경우 근거 수준이 낮아진다.

3) 근거의 간접성

두 가지 유형이 있다.

첫째, 대안 중재(A와 B)의 효과를 비교하는 문헌고찰에서 A와 위약 그리고 B와 위약을 비교했다. 따라서 근거는 중재 A와 B는 간접 비교했다고 할 수 있다. 간접적인 비교가 네트워크 메타분석 문맥 내에서 수행되는 경우, 네트워크 메타분석을 위한 GRADE가 사용되어야 한다.

둘째, 문헌고찰이 적격 기준에는 부합하지만, 인구, 중재, 비교, 중재결과 측면에서 문헌고찰 질문에 직접적이지 않은 무작위배정 임상시험일 수 있다. 예를 들어, 관상 동맥 심장 질환의 2차 예방을 위한 중재를 다루는 문헌고찰에서, 확인된 연구의 대부분은 당뇨병을 앓고 있는 사람들에게도 수행되었다고 하면, 인구집단이 주로 당뇨병 환자와 관련되기 때문에 간접적인 근거로 간주될 수 있다. 이와 반대되는 시나리오도 똑같이 적용될 수 있다. 당뇨병 환자의 관상 동맥 심장 질환에 대한 예방 전략 효과에 대한 문헌고찰을 가지고 당뇨병이 없는 사람들에 대한 간접적인 관련 근거를 제공할 수 있다.

간접성의 다른 원인은 중재에서 비롯될 수 있다. 예를 들어, 모든 포함된 연구에서 기술적인 중재가 전문 센터에서 숙련되고 고도로 숙련된 전문가에 의해 수행된 경우, 이들 센터 외부에서 수행된 중재 효과에 대한 근거는 간접적 일 수 있다. 비교 집단의 예로는 비교 집단이 대부분의 세팅에서 표준 치료보다 덜 효과적인 중재를 받은 경우가 그렇다. 중재결과 평가의 예로는 환자의 중요한 중재결과에 대한 자료가 없을 때 또는 조사자가 삶의 질에 대한 자료가 필요

하지만 증상만 보고될 때, 이러한 대리 결과로 인한 간접적인 경우가 해당된다. 문헌고찰 저자는 주요 관심 그룹에서 예상되는 효과의 차이를 기반으로 하향 평가가 정당하다고 판단될 때, 판단을 투명하게 해야 한다. 문헌고찰 저자는 GRADEpro GDT 소프트웨어에서 제공되는 표 15-2를 사용하면 간접성에 대한 판단의 투명성을 높일 수 있다.

Outcome: …					
Domain (original question asked)	Description (evidence found and included, including evidence from other studies) – consider the domains of study design and study limitation, inconsistency, imprecision and publication bias	Judgement – is the evidence sufficiently direct?			
Population:		Yes ☐	Probably yes ☐	Probably no ☐	No ☐
Intervention:		Yes ☐	Probably yes ☐	Probably no ☐	No ☐
Comparator:		Yes ☐	Probably yes ☐	Probably no ☐	No ☐
Direct comparison:		Yes ☐	Probably yes ☐	Probably no ☐	No ☐
Outcome:		Yes ☐	Probably yes ☐	Probably no ☐	No ☐
Final judgement about indirectness across domains:		☐ No indirectness	☐ Serious indirectness	☐ Very serious indirectness	

그림 15-2. 중재결과에 따른 간접성에 대한 판단(GRADEpro GDT에서 이용가능함)

4) 연구결과의 비정밀성

연구대상자가 거의 없거나 사건이 적고 신뢰 구간이 넓은 경우, 근거 수준을 낮출 수 있다. '결과 요약표'에 포함된 신뢰 구간을 통해 정밀도를 평가할 수 있는 정보를 얻을 수 있다. 표본 크기 계산과 유사하게 최적의 정보 크기(OIS: optimal information size) 또는 문헌고찰 정보 크기(RIS: review information size)의 계산을 사용하여 비정밀성을 판단할 수 있다. OIS는 적절한 연구의 검정력을 위해 개별 연구에 필요한 연구대상자 수를 기준으로 계산된다. 95% 신뢰 구간이 위험비(RR) 1을 통과하지 않고, 총 사건 또는 연구대상자수가 OIS 기준을 초과하면 정밀성이 적절하다고 판단한다. 95% CI가 상당한 이득이나 위해를 포함한다면(RR〈0.75 또는 RR〉1.25), OIS 기준이 충족되더라도 비정밀성에서 하향 평가되는 것이 적절할 수 있다.

5) 출판 비뚤림 가능성이 매우 높음

연구결과의 근거에 따라(일반적으로 효과가 없는 것에 대해 : 출판 비뚤림) 또는 중재결과에 따라(일반적으로 유해하거나 효과가 없을 수 있는 것에 대해 : 선택적 중재결과 비보고 비뚤림)연구를 보고하였으면 근거 수준이 낮아질 수 있다. 선택적 중재결과 보고는 비뚤림위험 평가의 일부로 연구 단계에서 평가되므로, 결과 요약표'의 결과에 기여하는 연구에서 영역 1(디

자인 및 실행의 제한점)에서 다루어진다. 문헌고찰에 포함된 다수의 연구가 중재결과에 기여하지 않거나 출판 비뚤림의 근거가 있는 경우 근거 수준은 낮아질 수 있다. 출판 비뚤림을 의심할만한 상황은 출판된 근거가 대부분 업계에서 자금을 지원한 소규모 연구에서 나올 때이다. 예를 들어, 치질 환자에서 플라바노이드에 대한 14가지 연구는 명백한 큰 이점을 보여 주었지만 총 1432명의 환자를 등록했다. 각 연구는 비교적 적은 수의 환자를 등록했다. 이러한 연구의 대부분에서 후원이 이루어졌기 때문에 이득이 없는 미출판 연구가 있을 거라는 의심이 생길 수 있다(Alonso-Coello et al 2006).

15.1.3 전체적 근거 수준을 높일 수 있는 영역

비록 NRSI와 하향 평가된 무작위배정 임상시험 때문에 근거 수준이 낮아질 수 있지만, 근거를 중등도 또는 심지어 높은 확신도로 '상향평가'할 수 있는 특별한 상황이 있다(표 15-1).

1) 큰 효과 : 드물지만 방법론적으로 잘 수행된 관찰 연구에서 중재 효과가 크고 일관적이면, 해당 결과에 대해 확신할 수 있다. 그럴듯한 교란이 없으면서 큰 추정 효과(예 : RR 〉 2 또는 RR 〈0.5), 또는 타당도를 위협하는 것이 없는 연구에서 매우 큰 효과(예 : RR〉 5 또는 RR 〈0.2)가 있을 수 있다. 이러한 상황에서, NRSI가 실제 효과를 과대 추정했다고 할지라도, 약한 연구설계만으로 관찰된 이득 모두를 설명하지 못할 수 있다. 따라서 관찰 연구설계에도 불구하고, 그 효과가 있음을 확신할 수 있다. 이러한 연구에서의 효과의 크기는(다른 방법론적 제한점이 없을 때 효과가 크다면) 낮은 수준에서 중간 수준으로 근거 수준을 바꿀 수 있다. 예를 들어 관찰 연구에 대한 메타분석에 따르면 자전거 헬멧은 자전거 타는 사람의 두부 손상의 위험을 큰 폭으로 줄였다(오즈비 0.31, 95% CI 0.26-0.37). 이러한 큰 효과는 명백한 비뚤림이 없을 때 중등도 근거 수준으로 등급을 제시한다(Thompson et al 2000).

GRADE 지침은 상대 효과가 2.0보다 큰 경우, 한 단계 근거 수준을 높게 평가할 것을 제안한다. 그러나 상대 효과의 점 추정치가 2.0보다 크지만 신뢰 구간이 2.0보다 상당히 작은 것을 포함한다면 큰 효과에 대해 평가를 내리는 결정이 어려울 수 있다.

2) 용량 반응 : 용량-반응 변화의 존재는 관찰 연구의 결과에 대한 신뢰를 증가시킬 수 있고 근거 수준을 높일 수 있다. 예를 들어, 치료 전 항응고제를 복용한 환자에서 출혈의 위험이 증가한다는 관찰 연구결과에 대해, 혈액 응고에 필요한 시간 사이과 출혈이 위험 사이의 '용량 - 반응' 변화가 있다는 관찰에 의해 항응고에 필요한 시간과 출혈 위험성의 증가를 확신하였다.

3) 그럴듯한 교란 : 때로는 무작위 연구나 비무작위 연구의 모든 그럴듯한 비뚤림이 명백히 중재 효과를 과소평가할 수 있다. 관찰 연구가 연관성을 증명하지 못했지만 모든 그럴듯한 비뚤림이 중재 효과를 증가시켰을 때도 유사 상황이 된다.

15.1.4 GRADE 틀을 사용하여 근거 수준에 대한 평가 내용을 기술

문헌고찰 저자는 하향 평가하거나 상향 평가 한 근거를 제시하고, '결과 요약표'를 참조하여, 결과 섹션에서 근거 수준에 대한 등급을 보고해야 한다.

표 15-1는 문헌고찰 저자가 각 도메인에서 근거 수준에 대한 판단을 어떻게 정당화할 수 있는지에 대한 틀과 예제를 제공한다.

표 15-1. 근거 수준 기술의 틀과 하향 평가 또는 상향평가에 대한 정당화

근거 수준 평가 영역	결과 섹션	근거 수준을 하향 평가 또는 상향 평가한 예시
비뚤림위험	비뚤림위험 표에서 사용된 기준에 기반 해서 기술한다.	10개의 무작위배정 임상시험 중 5개가 환자나 보호자 눈가림이 되지 않았기 때문에 하향평가 됨
비일관성	하나 이상의 지표(예. I^2와 p값)를 사용한 결과, 신뢰구간 겹침, 상이한 추정치, 연구간 변이에 의한 비일관성 정도를 기술한다.	이질성에 의한 효과추정치의 변동 비율이 낮아 하향평가 하지 않음(I^2=0%)
간접성	대부분의 연구가 PICO를 다루고 있는지 설명한다. 제시된 질문과 유사한가?	포함된 연구가 말기암 환자로 제한 되었기 때문에 하향평가 됨
비정밀성	사건 수와 신뢰 구간의 폭을 설명한다.	사망에 대한 효과의 신뢰 구간은 상당한 이득과 상당한 위해 모두에 걸쳐있어서 하향 평가 됨
출판 비뚤림	출판 비뚤림의 가능성 정도를 기술한다.	·14건의 무작위배정 시험의 깔때기 그림에서 표본 크기가 작고, 긍정적인 효과를 보이는 몇 가지 작은 연구가 있었지만, 효과가 없는 연구는 발표되지 않았을 수 있기 때문에 근거 수준이 낮아짐 · 3개의 규모가 작고, 긍정적인 연구가 있지만 효과나 위해가 없는 연구는 아직 발표되지 않은 것으로 보임. 또한 이해상충이 개입했을 가능성이 큼. 근거 수준이 낮아짐
근거 수준 평가 영역	**결과 섹션**	**근거 수준을 하향 평가 또는 상향 평가한 예시**
큰 효과 (상향 평가)	효과의 크기와 관련 신뢰 구간의 폭을 설명한다.	RR이 크므로 상향 평가됨 : 0.3(95 % CI 0.2 - 0.4), 충분한 사건 수
용량 반응 (상향 평가)	연구는 노출 수준이 높은 결과(예 : 폐암)에서 명백히 증가된 연관성을 보여준다.	흡연하지 않은 자에서 10%, 1년 10팩 흡연자에서 15%, 15팩 흡연자에서 20%의 상대적 위험 증가를 보여 용량-반응 관계가 나타나기 때문에 상향 평가되었다.

반대되는 그럴듯한 잠재적 비뚤림과 교란(상향 평가)	고려되지 않은 반대되는 비뚤림과 교란을 설명한다.	효과 추정치는 흡연, 교육 정도와 같은 흡연에 대한 교란에 대해서 통제되지 않았지만, 이러한 요소는 오히려 실제 효과를 과소평가하게 될 가능성이 있다. 근거 수준이 상향되었다.

15.2 결론 도출

15.2.1 코크란 리뷰의 결론 섹션

코크란 리뷰에서 저자의 결론은 연구가 임상진료에 미치는 영향, 연구에 미치는 영향으로 나누어져 있다. 중재에 대해서는 코크란 리뷰가 임상진료에서 의미 있는 정보와 지침을 제공할 수 있지만 여러 가지 의료 옵션에 대한 권고여부는 코크란 리뷰만으로는 부족할 수 있다. 임상진료에 미치는 영향이나 진료지침을 개발하기 위해 근거 수준, 이익과 위해의 균형, 가치와 선호도를 고려할 수 있다.

15.2.2 임상진료에 대한 영향

중재의 실질적인 유용성에 대한 결론을 도출하기 위해서는 예상되는 이익, 손해 및 가치와 선호를 고려해야 한다. 특정 상황에서 특정 행동에 대한 구체적인 권고를 하는 것은 코크란 리뷰를 넘어서는 것이며 코크란 리뷰가 제공하지 않는 추가 근거 및 정보에 근거한 판단이 필요하다. 이러한 판단은 일반적으로 임상 진료 지침 개발자의 영역이다. 따라서 코크란 리뷰에서는 권고문을 제시해서는 안 된다.

문헌고찰 저자가 임상의와 환자가 취할 수 있는 행동을 제시해야 한다면, 근거 수준과 이득과 위해의 균형을 기술 한 후에 특정 가치 패턴과 선호도와 일치하는 다양한 행동을 강조해야 한다.

중재의 영향, 환자의 기저 위험 또는 상태, 비용 및 그 비용을 지불하는 사람, 그리고 자원의 이용 가능성과 같은 효과를 수정할 것으로 기대되는 요소를 포함하여 의사결정에 영향하는 다른 요소도 강조되어야 한다.

문헌고찰 저자는 제한된 자료가 제공하는 환자 중재결과를 포함하여 중요한 모든 중재결과를 고려해야 한다. 공중보건 문헌고찰의 맥락에서, 대상 인구집단은 질병이 없는 인구일 수도 있고, 중재를 받는 집단에서 측정되지 않은 중재결과를 포함할 수도 있다. 이 과정은 다른 결과에 관련된 가치 또는 선호도와 근거 수준에 대한 판단에서 높은 수준의 설명을 포함한다.

15.2.3 연구에 대한 영향

저자의 결론에 대한 두 번째 범주는 연구에 대한 영향이다. 미래의 의료 연구에 대한 정보에 입각한 의사결정을 내릴 수 있도록 '연구에 대한 영향' 섹션에는 향후 연구가 필요하다는 점과 가장 바람직한 향후 연구의 방향이 무엇인지 언급해야 한다. 현재 문헌고찰 근거 수준에서 다루어지거나 보다 효과적으로 다루어 질 수 있는 인구, 중재, 비교 및 결과를 고려하는 것이 도움이 된다.

P(인구집단) : 진단, 질병 단계, 병적 상태, 위험 요소, 성별, 연령, 인종별, 구체적 포함 또는 제외 기준, 임상 세팅
I (중재) : 유형, 빈도, 용량, 기간, 예후 인자
C(비교) : 위약, 일상 관리, 대체 치료/관리
O(결과) : 어떤 임상적 또는 환자 관련 중재결과를 포함할지, 측정 방법, 개선 필요성 등

GRADE 도메인은 어떤 추가 연구가 근거 수준을 향상시킬 수 있는지 이해하고 설명할 수 있게 한다. 근거 수준은 중재결과마다 다양하므로, 이러한 중재의 영향은 특정 중재결과들에 따라 다를 것이다. 표 15-2는 문헌고찰 저자가 근거 자료를 해석하고 향후 연구 및 실행에 대한 결론을 도출하는 방법을 보여준다.

표 15-2. 개별 GRADE 도메인에서 제안한 연구 및 임상진료에 대한 영향

영역	연구에 대한 영향	연구 보고서의 예	임상진료에 대한 영향
비뚤림 위험	방법론적으로 더 잘 설계되고 실행된 연구가 필요하다.	모든 연구는 결과 평가자 눈가림이 부족했다. 이 유형의 임상시험이 필요하다.	중재결과에 대한 평가자 눈가림 부족 때문에 효과의 추정치가 편향될 수 있다.
비일관성	설명되지 않은 비일관성 : 개별 연구대상자 자료의 메타분석의 필요성; 관련 하위 그룹의 연구 필요.	췌장암 환자와 효과가 다른지 알아보기 위해서는 소세포 폐암 환자에 대한 연구가 필요하다.	설명되지 않은 비일관성 : 전체적 근거 수준에 대한 전반적인 효과 평가를 고려하고 해석한다. 설명된 비일관성(결과가 계층에 따라 제시되지 않은 경우) : 하위 그룹별 영향 평가를 고려하고 해석
간접성	관심 있는 PICO 질문에 더 잘 맞는 연구가 필요.	암이 진행된 환자의 연구결과이기 때문에 조기암 환자에 대한 연구가 필요하다.	결과가 환자에게 직접 적용되는지 또는 특정 환경에서 중재가 적용되는 방식인지는 확실하지 않다.

영역	연구에 대한 영향	연구 보고서의 예	임상진료에 대한 영향
비정밀성	최적의 정보 크기에 도달하기 위해 더 많은 연구대상자와 더 많은 연구가 필요	실험 중재군과 비교 중재군에서 약 200 건의 사건에 대한 연구가 필요하다.	전체적 근거 수준으로 동일한 불확실성 해석 : 예. 진정한 효과는 실질적으로 다를 수 있다.
출판 비뚤림	게시되지 않은 자료를 조사하고 확인해야 한다. 대규모 연구가 이 문제를 해결하는 데 도움이 될 수 있다	대규모 연구가 요구된다.	전체적 근거 수준으로 동일한 불확실성 해석 : 예. 진정한 효과는 실질적으로 다를 수 있다.
큰 효과	직접적인 영향이 없다.	적용할 수 없음	효과는 연구에 포함된 집단에서 크며, 진정한 효과는 중요한 임계치를 넘을 가능성이 크다.
용량 효과	직접적인 영향이 없다.	적용할 수 없음	노출이 감소 될수록 위해(또는 이득)이 더 커질 것이다.
반대되는 비뚤림과 교란	잠재적 비뚤림과 교란변수를 통제할 연구가 필요하다.	흡연과 교육 정도와 같은 가능한 교란변수를 통제하는 연구가 필요하다.	효과는 제시된 것보다 크거나 작아질 수 있다.

항공기 탑승자의 심부 정맥 혈전증 예방을 위한 압축 스타킹의 검토에 대한 체계적 문헌고찰(Clarke et al 2016)은 중재의 이점에 대한 확실한 근거에 대한 예이다. 이 문헌고찰은 DVT 무증상 효과에 대해 압축 스타킹을 착용한 것과 하지 않은 것을 비교하였다. 이런 문헌고찰을 통해 다른 예방 전략과 비교한 스타킹 중재의 효과나 스타킹의 다른 강점에 대한 추후 연구를 정당화할 수 있다. 사망, 폐색전증 및 증상이 있는 DVT와 같은 중재결과에 대해, 압축 스타킹을 착용하지 않는 것과 착용 할 때의 효과에 대한 나머지 불확실성을 해결하기 위한 무작위배정 임상시험은 더 할 필요가 있다.

불안 장애를 위한 치료적 접촉의 문헌고찰(Robinson et al 2007)은 적격한 연구가 발견되지 않은 상황에서 연구에 대한 영향을 어떻게 기술한 것인지의 예를 제공한다. 즉 이 문헌고찰에서는 불안으로 진단받은 사람들의 불안 증상 감소에 대한 치료적 접촉의 효과를 평가하기 위한 무작위배정 비교연구의 필요성을 강조한다. 향후 임상시험은 중재결과의 평가 및 해석을 가능하게 하는 방법론의 모든 측면에 대한 양질의 설명을 포함하여 방법론적으로 엄격할 필요가 있다.

15.2.4 결론 도달하기

'효과의 근거 없음'과 '효과가 없는 근거'를 혼동해서는 안 된다. 신뢰 구간이 너무 넓은 경우(예 : 효과가 없는 경우를 포함해서) 실험 중재가 '효과 없음' 또는 '차이 없음'으로 결론

내리면 안 된다.

일부 효과를 '긍정적으로' 잘못 결론 내릴 수 있다. 예를 들어, 효과 추정에서 이득 중재결과에 대해 긍정적이지만 신뢰 구간이 넓은 경우, 문헌고찰 저자는 효과가 있다고 기술 하면 안 된다. 또한 결과에 대한 효과 추정치가 부정적이지만 신뢰 구간은 '효과 없음'을 포함할 때도, 효과 없음으로 보고해서는 안 된다.

원하는 방향으로 결론을 내는 것도 잘못된 것이다. 예를 들어 포함된 연구에서 사망률의 감소 또는 증가가 통계적으로 유의하지 않은 것으로 나타났을 때 "사망률 감소를 확인하는 데 분석된 사람 수가 너무 적었다"라고 쓰는 것은 잘 못된 것일 수 있다.

이러한 오류를 피하는 한 가지 방법은 연구결과가 눈가림 되었다고 간주하는 것이다. 즉, 연구결과의 방향이 바뀌면 결과가 어떻게 제시되고 결론이 내려지는지를 고려하면 된다. 중재의 효과 차이의 추정치에 대한 신뢰 구간이 '효과 없음'과 겹치면, 분석은 진정한 유익한 효과와 진정한 해로운 효과 모두 양립할 수 있다. 결론 중 하나에 가능성이 언급되면 다른 가능성도 언급되어야 한다. 표 15-3은 메타분석에서 얻은 효과 추정치와 근거 수준에 근거하여 결론을 내기 위한 서술문을 제안한다.

표 15-3. 결론을 위한 제안된 서술문

효과 추정치에 따른 효과크기	결론을 위해 제안된 진술문 (X는 중재를 대신한다, 효과 방향에 따라 '감소나 '증가'를 선택한다, "중재결과' 대신 중재결과 이름을 대신한다, 필요시 'Y와 비교할 때'를 포함한다.)
높은 근거 수준	
큰 효과	X는 중재결과에서 큰 감소/증가를 가져온다.
중등도 효과	X는 중재결과를 감소/증가시킨다. X는 중재결과에서 감소/증가를 가져온다.
작은 중요한 효과	X는 중재결과를 약간 감소/증가시킨다. X는 중재결과에서 약간의 감소/증가를 가져온다.
사소한, 작은, 중요하지 않은 효과 혹은 효과 없음	X는 중재결과에서 매우 적은 차이를 가져온다. X는 중재결과를 감소/증가시키지 않는다.
중등도 근거 수준	
큰 효과	X는 아마도 중재결과에서 큰 감소/증가를 가져온다. X는 중재결과에서 큰 감소/증가를 가져올 가능성이 있다.

중등도 효과	X는 아마도 중재결과를 감소/증가시킨다. X는 중재결과에서 감소/증가를 가져올 가능성이 있다.
작은 중요한 효과	X는 아마도 중재결과를 약간 감소/증가시킨다. X는 중재결과에서 약간의 감소/증가를 가져올 가능성이 있다.
사소한, 작은, 중요하지 않은 효과 혹은 효과 없음	X는 아마도 중재결과에서 매우 적은 차이를 가져온다. X는 중재결과에서 차이를 가져오지 않을 가능성이 있다. X는 아마도 중재결과를 감소/증가시키지 않는다. X는 중재결과를 감소/증가시키지 않을 가능성이 있다.
낮은 근거 수준	
큰 효과	X는 중재결과에서 큰 감소/증가를 가져올 수 있다. 근거는 X가 중재결과의 큰 감소/증가를 가져온 것을 시사한다.
중등도 효과	X는 중재결과를 감소/증가시킬 수 있다. 근거는 X가 중재결과를 감소/증가시켰다는 것을 시사한다. X는 중재결과에서 감소/증가를 가져온다. 근거는 X가 중재결과의 감소/증가를 가져온 것을 시사한다.
작은 중요한 효과	X는 중재결과를 약간 감소/증가시킬 수 있다. 근거는 X가 중재결과를 약간 감소/증가시켰다는 것을 시사한다. X는 중재결과에서 약간의 감소/증가를 가져온다. 근거는 X가 중재결과의 약간의 감소/증가를 가져온 것을 시사한다.
사소한, 작은, 중요하지 않은 효과 혹은 효과 없음	X는 중재결과에서 매우 적은 차이를 가져올 수 있다. 근거는 X가 중재결과의 매우 적은 차이를 가져온 것을 시사한다. X는 중재결과를 감소/증가시키지 않을 것이다. 근거는 X가 중재결과를 감소/증가시키지 않는다는 것을 시사한다.
매우 낮은 근거 수준	
어떤 효과	중재결과에 대한 X의 효과에 대한 근거는 매우 불확실하다. X는 중재결과를 감소시키거나 증가시키거나 거의 영향을 미치지 않지만 그 근거는 매우 불확실하다.

15.3 근거 수준 평가와 결론 도출 관련 MECIR

C74 : 전체적 근거 수준 평가(필수)	
각 결과에 대한 전체적 근거 수준을 평가하고, 문헌고찰 본문에 있는 근거의 수준에 관한 결론을 도출하기 위한 다섯 가지 GRADE 고려사항(비뚤림위험, 효과의 일관성, 비정밀성, 간접성, 또는 출판 비뚤림)을 이용한다.	GRADE는 여러 연구의 중재 효과에 대한 신뢰도를 요약하는 데 가장 널리 사용되는 방법이다. 온라인 GRADEpro 도구를 사용하고 소프트웨어의 도움말 시스템에 설명된 대로 사용하는 것이 좋다. 이렇게 하면 저자자 팀은 자신의 판단을 알리기 위해 같은 정보에 접근할 수 있다. 이상적으로, 독립적으로 두 사람이 전체적 근거 수준을 평가하고, 등급을 낮추는 결정에 대한 의견 일치를 이끌어 내야한다. 다섯가지 GRADE 고려 사항은 문헌고찰이 '결과 요약표'를 포함했는지 여부에 관계없이 다루어져야 한다. 이 정보는 논의나 저자의 결론 작성에 도움이 되고, 초록 및 Plain language 요약에서 근거 수준을 기술하는 데 도움이 된다.
C75 : 전체적 근거 수준 평가에 대한 근거 제시하기(필수)	
전체적 근거 수준에 대한 모든 평가에 대한근거를 제시하고 문서화한다(예 : GRADE를 사용하여 하향평가 또는 상향평가에 대한 것).	구조화된 접근법을 채택하면 근거의 해석을 공식화할 때 투명성이 보장되고 그 결과는 사용자에게 보다 유익하다.

15.4 참고문헌과 읽을거리

- Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S, Guyatt GH, Harbour RT, Haugh MC, Henry D, Hill S, Jaeschke R, Leng G, Liberati A, Magrini N, Mason J, Middleton P, Mrukowicz J, O'Connell D, Oxman AD, Phillips B, Schünemann HJ, Edejer TT, Varonen H, Vist GE, Williams JW, Jr., Zaza S. Grading quality of evidence and strength of recommendations. BMJ 2004;328: 1490.
- Clarke MJ, Broderick C, Hopewell S, Juszczak E, Eisinga A. Compression stockings for preventing deep vein thrombosis in airline passengers. Cochrane Database of Systematic Reviews 2016; 9: CD004002.
- Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, Norris S, Falck-Ytter Y, Glasziou P, DeBeer H, Jaeschke R, Rind D, Meerpohl J, Dahm P, Schünemann HJ. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. Journal of Clinical Epidemiology 2011a;64: 383-394.

- Guyatt GH, Oxman AD, Kunz R, Brozek J, Alonso-Coello P, Rind D, Devereaux PJ, Montori VM, Freyschuss B, Vist G, Jaeschke R, Williams JW, Jr., Murad MH, Sinclair D, Falck-Ytter Y, Meerpohl J, Whittington C, Thorlund K, Andrews J, Schünemann HJ. GRADE guidelines 6. Rating the quality of evidence--imprecision. Journal of Clinical Epidemiology 2011b;64: 1283-1293.
- Robinson J, Biley FC, Dolk H. Therapeutic touch for anxiety disorders. Cochrane Database of Systematic Reviews 2007; 3: CD006240.
- Schünemann HJ, Higgins JPT, Vist GE, Glasziou P, Akl EA, Skoetz N, Guyatt GH. Chapter 14: Completing 'Summary of findings' tables and grading the certainty of the evidence. Draft version(29 January 2019) for inclusion in: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA(editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.0(updated July 2019). Cochrane, 2019. Available from www.training.cochrane.org/handbook.
- Schünemann HJ, Vist GE, Higgins JPT, Santesso N, Deeks JJ, Glasziou P, Akl EA, Guyatt GH. Chapter 15: Interpreting results and drawing conclusions. Draft version(29 January 2019) for inclusion in: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA(editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.0(updated July 2019). Cochrane, 2019. Available from www.training.cochrane.org/handbook.

PART II

체계적 문헌고찰 방법론의 주요 이슈



목차 PART II. 체계적 문헌고찰 방법론의 주요 이슈

제16장 형평성

16.1 체계적 문헌고찰에서 형평성 : 개요 ····· 227
16.2 문헌고찰 구성하기(formulation) ····· 228
16.3 근거 검색(identification) ····· 231
16.4 근거 평가 ····· 232
16.5 근거 합성(synthesis) ····· 232
16.6 근거 해석 ····· 233
16.7 맺음말 ····· 233
16.8 참고문헌과 읽을거리 ····· 234

제17장 중재의 복합성

17.1 개요 ····· 235
17.2 문헌고찰의 구조화(formulation) ····· 240
17.3 근거의 확인 ····· 243
17.4 근거 평가 ····· 244
17.5 근거의 합성 ····· 244
17.6 근거의 해석 ····· 245
17.7 참고문헌과 읽을거리 ····· 246

제18장 개별 연구대상자 자료 분석

18.1 소개 ····· 247
18.2 IPD 수집 ····· 250
18.3 IPD 관리 및 확인 ····· 253
18.4 IPD 분석 ····· 254
18.5 IPD 문헌고찰의 품질 평가 ····· 255
18.6 참고문헌과 읽을거리 ····· 255

제19장 기존의 체계적 문헌고찰의 활용 (Using Existing Systematic Reviews)

19.1 문헌검색 ····· 259
19.2 체계적 문헌고찰의 질평가 ····· 259
19.3 자료추출 ····· 261
19.4 추가검색 ····· 262
19.5 기존 체계적 문헌고찰 이용 방법 ····· 262
19.6 방법과 결과의 보고 ····· 262
19.7 관련 이슈 ····· 263
19.8 참고문헌과 읽을거리 ····· 264

제20장 신속 체계적 문헌고찰

20.1 개요 ····· 265
20.2 신속 체계적 문헌고찰의 과정 ····· 267
20.3 참고문헌과 읽을거리 ····· 277

제21장 경제성 근거

21.1 경제성 근거 문헌고찰 방법에 대한 핵심 원칙 ····· 279
21.2 문헌고찰에서 경제성 근거를 포함할지에 대한 우선순위를 정하는 기준 ··· 280
21.3 간단한 경제성 논평 과정 ····· 281
21.4 참고문헌과 읽을거리 ····· 287

제22장 질적 근거 합성

22.1 중재고찰과 연계된 질적 근거 합성 계획 ····· 290
22.2 질문 개발 ····· 291
22.3 질적 근거 검색 ····· 292
22.4 질적 연구의 방법론적 강점과 제한점 평가 ····· 293
22.5 문헌 선택 ····· 295
22.6 질적 근거 합성법 선택 ····· 295
22.7 자료추출 ····· 296
22.8 질적 합성 결과에 근거 수준 평가 ····· 298
22.9 질적 근거 합성을 중재문헌고찰에 통합하는 방법 ····· 298
22.10 질적 근거 합성 및 프로토콜 보고 ····· 299
22.11 참고문헌과 읽을거리 ····· 300

제23장 특수 무작위배정 임상시험 포함

23.1 군집 무작위배정 임상시험 ····· 302
23.2 교차 임상시험(Corssover trials) ····· 307
23.3 참고문헌과 읽을거리 ····· 312

제24장 유해효과

24.1 유해효과에 대한 접근법 ····· 315
24.2 연구설계 유형 선택 ····· 317
24.3 관심 유해효과 선택 ····· 317
24.4 근거 검색 ····· 318
24.5 유해효과에 대한 비뚤림위험 평가 ····· 318
24.6 근거의 합성과 해석 ····· 319
24.7 참고문헌과 읽을거리 ····· 320

목차 PART II. 체계적 문헌고찰 방법론의 주요 이슈

제25장 근거 축적을 위한 전향적 접근

25.1 근거 감시: 근거 축적에 대한 능동적 감시 ····· 321
25.2 전향적으로 계획된 메타분석 (Prospectively planned meta-analysis) 325
25.3 참고문헌과 읽을거리 ····· 330

제26장 문헌고찰 Overview

26.1 서론 ····· 331
26.2 Overview란 무엇인가? ····· 331
26.3 Overview는 언제 시행해야 하는가? ····· 334
26.4 Overview 수행을 위한 방법 ····· 336
26.5 참고문헌과 읽을거리 ····· 347

제27장환자보고 중재결과 (Patient-reported outcomes)

27.1 PRO 소개 ····· 349
27.2 문헌고찰을 구성하기 ····· 350
27.3 근거의 평가(appraisal) ····· 352
27.4 합성 및 근거의 해석 ····· 354
27.5 참고문헌과 읽을거리 ····· 356

표 목차

표 19-1. 기존의 체계적 문헌고찰 활용 ····· 257
표 19-2. 선택된 체계적 문헌고찰에서 이용 가능한 자료 목록.(AHRQ, 2009) Adapted table AHRQ(Original publisher) ····· 262
표 20-1 . 신속 체계적문헌고찰에서 가능한 방법과 타당도에 미치는 영향····· 266
표 20-2. 신속 체계적 문헌고찰 핵심 질문 예시 ····· 268
표 20-3. 신속 체계적 문헌고찰 프로토콜의 내용 ····· 269
표 20-4. 신속 체계적 문헌고찰의 검색 전략····· 270
표 20-5 . 기존 체계적 문헌고찰의 활용····· 273
표 20-6. 신속리뷰의 검색전략 ····· 274
표 20-7. 신속 체계적 문헌고찰의 보고서 목차 ····· 275
표 20-8. 신속 체계적 문헌고찰 수행을 정보 기술(Valerie et al., 2017)····· 276
표 21-1. 경제적 근거의 포함에 대한 문헌고찰의 우선순위를 정하는데 도움이 되는 결정 알고리즘····· 281
표 21-2. MEDLINE(OvidSP) filter for cost-of-illness studies ····· 283
표 21-3. EMBASE(OvidSP) filter for cost-of-illness studies ····· 283
표 21-4. PsycINFO filter for cost-of-illness studies····· 284
표 21-5. 경제성 논평 적격기준 작성의 예시 ····· 284
표 22-1. 질적 근거 합성을 위한 PerSPecTIF 질문형식 틀 (Booth et al (2019b) Reproduced with permission of BMJ Publishing Group ····· 292
표 22-2. 중재문헌고찰이나 혼합방법 문헌고찰로 통합하기 위한 질적 근거 합성 수행시 권고되는 방법론····· 295
표 22-3. '포함된 연구의 특성' 표에 포함될 맥락 및 방법론 정보 ····· 297
표 23-1. 군집 무작위배정 임상시험 ROB 도구 평가에서 다루어지는 이슈····· 304
표 23-2 . 교차 무작위배정 임상시험 ROB 도구 평가에서 다루어지는 이슈····· 310
표 26-1. Overview로 출판하기에 적당한 보건의료 중재에 대한 연구질문의 유형 ····· 333
표 26-2. Overview와 중재 문헌고찰 사이의 방법 비교 ····· 335
표 26-3. 포함된 체계적 문헌고찰에 포함된 주요 연구를 도식화하는 표 템플릿* 341
표 26-4. Overview 저자가 포함된 체계적 문헌고찰에서 추출하고자 할 수 있는 체계적 문헌고찰(및 그 일차 연구)의 서술적 특성 ····· 344
표 26-5. '결과 요약표' 템플릿 ····· 346
표 27-1. 임상시험에서 PROM을 설명하고 평가하기 위한 체크리스트 (Guyatt et al, 1997)····· 351

그림 목차

그림 19-1. 기존의 체계적 문헌고찰을 활용한 의료기술평가 ····· 258
그림 19-2. 기존 체계적문헌고찰 활용 방법 결정을 위한 알고리듬(CPSTF). ···· 261
그림 19-3. 기존의 체계적 문헌고찰 이용 연구 흐름도 ····· 263
그림 20-1. 체계적 문헌고찰 필요성과 방법 결정 알고리즘 ····· 268
그림 20-2. 기존 체계적 문헌고찰의 이용 방법에 따른 신속 체계적 문헌고찰 방법 · 272

박스 목차

Box 21-1. 간단한 경제성 논평에서 간결한 논의를 위한 문장의 예 ····· 287
Box 22-1. 질적 근거 합성을 위한 적절한 방법을 선택하기 위한 RETREAT 고려사항 ····· 291
Box 22-2. 7S framework ····· 293
Box 22-3. 연구엄격성 평가를 위한 방법론적 강점과 제한점 평가 영역 ····· 294
Box 24-1. 특정 중재의 유해효과를 검토하기 위한 세 가지 접근법의 예 : 침술 · 316

제16장 형평성

핵심 포인트

- 건강 형평성(health equity)은 건강에서 회피가능한, 그리고 불공평한 차이가 없는 것을 의미한다.
- 건강 불평등(health inequity)은 PROGRESS-plus에 의해 정의된 특성에 따라 확인할 수 있다.(PROGRESS:Place of residence, Race/ethnicity/culture/language, Occupation, Gender/sex, Religion, Education, Socio-economic status, Social capital; plus : 성적 지향, 나이, 장애 등과 같은 다른 특성).
- 코크란 리뷰는 인구집단 내에서 효과의 분포와 형평성을 고려함으로써 의사결정에 필요한 정보를 줄 수 있다.
- 건강 형평성 문제를 해결하기 위한 방법으로 다음과 같은 것이 있다.
  - 연구질문을 만들 때 로직 모델을 고려한다.
  - 형평성 및 특정 집단과 관련된 근거를 확인하고 평가하기 위해 어떤 방법을 사용할 것인지 결정한다.
  - '결과 요약표'에 반영할 것을 고려한다.(예. 취약계층을 위한 별도의 표 사용, 사건에 대한 위험의 차이에 대한 별도의 표 사용 등)
  - 고찰 파트에서 건강 형평성과 관련한 결과를 해석한다.

16.1 체계적 문헌고찰에서 형평성 : 개요

건강 형평성은 사회 정의에 대한 관심을 반영한다. 건강 차이를 피할 수 있고, 해결될 수 있고, 부당하고 불공평한 것으로 간주될 때, 건강 불평등을 고려한다. 모든 건강 차이가 불평등한 것은 아니다. 예를 들어 겸상 적혈구 질환(sickle sell disease)은 유전적 차이로 인해 일부 민족에서 더 흔하게 나타나지만 불공평한 것으로 간주되지는 않는다. 그러나 대기 오염 물질 분포의 차이로 인해 나타나는 아동기 천식 비율은 불평등한 것으로 간주될 것이다.

취약(disadvantaged)'은 사회적, 정치적, 법적 구조와 과정에 의해 야기되는 불리함이다. 잠재적 취약 축은 PROGRESS-plus(Place of residence, Race/ethnicity/culture/language, Occupation, Gender/sex, Religion, Education, Socio-economic status, Social capital)라는 약어로 정의내릴 수 있다. 여기에서 'Plus'는 나이, 성적 지향 및 개인의 건강에 대한

기회에 영향을 줄 수 있는 장애와 같은 추가적인 범주를 의미한다(O'Neill et al 2013). 문헌고찰의 독자층에 따라 건강이나 SCRAP(Sex, Comorbidities, Race, Age, Pathophysiology)과 같은 다른 특성의 목록이 도움이 될 수 있다.

평균적인 연구결과는 서로 다른 인구집단 내에서 효과의 차이를 숨긴다. 따라서 문헌고찰 저자는 평균적으로 효과가 있다고 할뿐만 아니라 건강 불평등에 대한 중재의 영향도 고려해야 한다. 체계적인 문헌고찰은 세 가지의 유형의 중재에 따라 건강 형평성에 대한 영향을 평가할 수 있다(Welch et al 2012):

일반 인구집단을 대상으로 하는 중재이고, 하나 이상의 PROGRESS-Plus 특성에 대한 효과 분포를 이해하는 것이 중요함.

- 사회적 약자 또는 위험에 처한 인구집단을 대상으로 하는 중재로, 형평성 결과는 없지만 불평등 감소에 대한 근거를 제공할 수 있음.
- 인구집단 전체 또는 하위 집단 간에 사회적 기울기(social gradient)[2]를 감소시키는 것을 목적으로 하는 중재

임상시험은 종종 취약 계층이나 특정 연령보다 높거나 낮은 사람들을 제외한다. 이런 집단을 배제하는 것은 임상시험에서 설정한 세팅 이외에 결과를 적용하는데 영향을 줄 수 있다. 문헌고찰 저자는 해당 PROGRESS-plus 요인에 따라 보고할뿐만 아니라 어떤 상태나 문제의 위험이 더 큰 하위 집단이 있는지, 또는 의료 접근성에 영향을 미치는 요인에 차이가 있는지 여부를 보고해야 한다. 이러한 요인들에는 환자/대중의 가치, 선호도, 가격 수용성 및 실현 가능성, 의사에 의한 의식적 또는 무의식적 편견이 포함될 수 있다.

PROGRESS-plus의 모든 특성을 평가하는 것은 일반적으로 불가능하다. 따라서 평가할 특성을 선택할 때, 중재의 수혜자와 근거의 사용자의 관점을 고려해야 한다.

16.2 문헌고찰 구성하기(formulation)

문헌고찰 질문을 구성하는데는 다음의 다섯 가지 이슈가 중요하다. 1) 건강 형평성 정의, 2) 형평성 및 로직 모델과 관련된 가설, 3) 적절한 연구설계, 4) 적절한 중재결과, 5) 맥락(context)

16.2.1 건강 형평성 정의하기

건강 형평성이라는 개념에는 공정성에 대한 판단이 포함되어 있기 때문에, 문헌고찰 저자가

2) "건강에서의 사회적 기울기는 사회경제적 위치에 있어서 혜택을 덜 받는 사람들이 더 많은 혜택을 받는 사람들보다 건강에서 더 나쁜 현상을 묘사하기 위해 사용되는 용어이다."(Institute of Health Equity)

해야 할 첫번째 단계는 평가 대상인 상태/문제 혹은 중재와 관련하여 건강 불평등을 경험하는 인구를 정의하는 것이다. 예를 들어 학교 급식에 대한 코크란 리뷰에서 사회경제적 수준, 성별 및 농촌 거주는 건강 불평등과 연관성이 있었지만, 기저상태에서의 영양상태가 사회경제적 불이익의 대리 측정치로 사용되었다(Kristjansson et al 2007). 대리 측정치 사용에 대한 정당성이 주어져야 하고, 그 용도가 투명해야 하며, 연구질문이나 중재의 수혜자에 따라 넓은 범위 문헌고찰 또는 좁은 범위 문헌고찰을 통해 특정 인구집단을 개별적으로 고려할 필요가 있다. 다양한 인구집단에 전달되는 중재의 경우 문헌고찰 저자는 연구대상자 일차 연구들이 관련 인구 통계를 투명하게 보고하고 있는지 평가해야 한다.

또한 잠재적 영향의 차이를 탐색하기 위해 민감도 또는 하위군 분석의 필요성을 평가하는 것도 중요하다. 형평성 문헌고찰은 하나 이상의 PROGRESS-plus 요인(예. 이민자, 소수 언어 집단, 노숙자)에 의해 정의된 인구집단 간의 차이를 고려할 수 있다. 그러나 모든 PROGRESS-plus 요인을 다룰 수는 없다. 따라서 질문 구성 단계에서 문헌고찰 저자는 어떤 요소가 가장 중요한지와 문헌고찰 할 때 어떻게 해결할 것인지를 명확히 고려해야 한다.

16.2.2 형평성에 대한 가설을 분명하게 하기 위한 로직 모델과 변화 이론

체계적 문헌고찰에 로직 모델, 인과 관계망 및 감소의 깔때기(funnels of attrition)와 같은 분석툴이 점점 많이 사용되고 있다. 입력과 결과물 사이의 관계를 보여주는 로직 모델은 반응의 차이, 기저상태 위험의 차이, 적용 가능성 및 효과를 중재할 수 있는 요인을 예측하여 건강 형평성에 미치는 영향을 평가하는 것과 관련된 주요 질문을 확인하는데 도움이 될 수 있다. 로직모델을 통해 문헌고찰 질문의 범위를 정하고, 적격기준을 확인하고, 검색 전략을 수립하고, 정책 혹은 임상에 적절한지를 고려할 수 있다.

16.2.3 형평성 평가를 위한 적절한 연구설계

형평성 평가를 위한 적절한 연구설계를 선택하여야 하며, 이론적 근거가 명확하게 설명되어야 한다. 특히 비무작위 연구가 인구집단 및 환경에 대한 중재의 영향 연구에서 적절하고 의미있는 근거를 제공할 수 있는지 여부를 고려하는 것이 필요하다.

16.2.4 형평성에 대한 적절한 중재결과

중재결과는 이해 관계자나 사용자군이 누구인지에 따라 선택될 필요가 있다. 적절한 군을 정의하는데 특정 틀이 도움이 될 수 있다. 예를 들어, 9P라는 틀이 있는데 그것은 patients, practitioners, public, policy-makers, press, product-makers, payers(예. 건강 보험자), purchasers(예. 고용주, 정부), principal investigators이다. 고려되어야 할 중재결과는 이득

과 위해이다 :(일반적/질병 특이적) 사망률, 중재영향(증상, 신체적/감정적/사회적/영적 기능, 삶의 질, 효용, 불편함, 재정적 부담), 그리고 중간/대리 연구결과/바이오마커 등

건강 및 사회적 중재결과의 상대적 중요성은 건강 불평등을 경험하는 인구집단에 따라 다를 수 있다. 예를 들어, 취약 계층 어머니의 취학 전 아동을 위한 돌봄(day care[3])에 대한 코크란 리뷰에서 중요한 중재결과는 모성 고용, 가족 수입 및 교육이었다(Zoritch et al 2000). 이러한 중재결과는 사회 경제적 수준이 높은 산모에게는 덜 중요할 수 있다. 노인 연령에서도 상대적 중요성에 대한 비슷한 분석이 가능한데, 연금이나 다른 형태의 사회 보장 제도가 노인연령에서는 중요한 중재결과가 될 수 있다.

'결과 요약표' 에 들어갈 중재결과를 선택할 때 세팅과 인구집단에 따라 중재결과의 중요성에 따라 점수를 매길 필요가 있다. 맥락은 중재결과의 중요도의 점수를 매길 때 고려되어야 한다. 또한 불편함, 부담(예. out-of-pocket 비용, 여행 시간), 그리고 낙인(stigma) 등도 일차 연구에서 일반적으로 보고되지 않았더라도 중재 대상자에게 굉장히 중요할 수 있으므로 잠재적 중재결과로서 고려될 필요가 있다.

16.2.5 맥락(context)과 형평성

중재가 계획되고 실행되는 사회적, 문화적, 정치적 맥락을 고려해야 한다. 보건 사회적 중재에 대한 일차 연구는 특정한 시간적, 문화적, 지리적, 정치적, 조직적 세팅에서 이루어지며, 이는 중재의 효과에 영향을 줄 수 있다.

'맥락을 고려함'은 맥락의 중요한 측면들을 이해하는 것으로, 이것들이 중재(예. 실행)에 어떻게 영향을 미칠 수 있는지, 중재결과에 영향을 미치는 정도를 기술하고 계층화 및 탐색하는 것을 의미한다. 예를 들어, 노인을 포함하는 문헌고찰의 경우, 통합적인 진료를 받지 못하는 다질환(multimobodity)과 전반적인 능력의 감소는 중요한 맥락적 이슈이다. 한 가지 측면은 처방된 의약품의 개수로 평가할 수 있으므로 문헌고찰 저자는 이 지표가 보고되기를 원할 수 있다. Context and Implementation of Complex Interventions(CICI) 프레임워크를 포함하여 상황에 맞는 자료를 수집하고 추출하기 위한 몇 가지 도구들이 개발되어 있다(Pfadenhauer et al 2017).

포함된 연구의 연구절차가 국제 윤리적 기준을 충족시키는지 평가하고 문서화하는 것이 필요할 수 있다. 왜냐하면 건강 불평등을 겪고 있는 인구집단은 대부분 취약한 피험자이며, 추가적인 보호수단이 필요할 수 있기 때문이다. 체계적 문헌고찰에서 포함된 연구의 윤리적인 우려를 규명함으로써 윤리적 실천을 강화할 수 있다.

연구 간 맥락의 다양성은 질적으로, 혹은 양적으로 평가될 수 있다. 맥락은 일차 연구의 여러 부분이나 첨부된 논문, 보고서, 정책 또는 역사적 문서에 기술될 수 있는데, 이러한 기술들을 찾기 위해서는 전문적인 지식이 필요할 수 있다. 따라서 전체 팀과 자문위원회 또는 기타

3) Day care: 미취학 아동·고령자·환자를 낮 동안 집이 아닌 시설에서 맡아 보살펴 주는 것

핵심 이해관계자는 문헌고찰이 유용하고 타당하며 적용가능한지 확인해야 한다.

16.3 근거 검색(identification)

형평성에 초점이 맞추어진 문헌고찰에 있어 검색은 일반적인 지침을 따르면 되지만, 관심 있는 인구집단을 충분히 커버해야 하는 문제가 있다. 건강 형평성과 관련된 검색은 생의학적 검색어로 커버가 힘든 경우가 많다. 따라서 관련 연구는 광범위한 문헌 소스에서 발견될 수 있으며 문헌에 대한 분류 자체가 신뢰성이 없는 경우도 많다. 이런 사실은 선택한 데이터베이스와 검색 용어에 영향을 줄 수 있다.

16.3.1 고려할 데이터베이스

관심 있는 중재결과로, 예를 들어, 노동 생산성이나 교육적, 경제적, 사회적 연구결과 같은 것들이라면 비의학 데이터베이스가 적절할 수 있다. Campbell 연합의 정보검색 지침은 사회적 결과변수와 관련된 검색에서 훌륭한 소스이다. 반면에 EPOC(Effective Practice and Organisation of Care) 노르웨이 지역 그룹에서는 저소득, 중간소득 국가에 적절한 데이터베이스 리스트를 가지고 있다(EPOC, 2013). 예를 들어 말라리아 예방을 위해 살충제 처리된 그물의 소유권과 사용을 증가시키는 전략에 대한 코크란 EPOC의 문헌고찰은 MEDLINE, EMBASE 외에 CINAHL, Web of Science, 학위논문, African Index Medicus, LILACS, WHOLIS 를 추가로 검색하였다(Augustincic Polec et al 2015).

16.3.2 형평성에 대한 검색어 선정과 검색 필터의 사용

PROGRESS-plus 범주를 설명하는 많은 단어가 주요 데이터베이스에서 색인되지 않기 때문에 표준 검색 필터(즉, 데이터베이스의 검색 인터페이스에서 사용 가능한 필터)를 사용하는 것은 상당한 위험을 동반한다(MEDLINE/Pubmed는 2016년에 '건강 형평성'을 MeSH에 새로 추가했다). 소아 연구들은 종종 잘 색인되지 않는다. 소아의 특정조건에 대한 연구 저자는 제목, 초록 또는 원고 내에서도 소아 용어를 명시적으로 사용하지 않을 수도 있다. 따라서 데이터베이스를 검색할 때 색인 또는 연령 제한만 사용하는 대신 소아 검색 필터(MeSH 용어, 연령 MeSH, 프리 텍스트 단어의 조합)를 사용하는 것이 좋다. 노인과 관련된 연구를 검색하는 것은 타당성을 위해 사용 가능한 검색 필터를 고려할 수 있다. 검증된 필터를 사용할 수 있으면 검색식을 짜고 스크리닝하는 문헌의 수를 줄이는 데 시간을 절약할 수 있다.

16.3.3 실용적인 조언

적절한 검색 전략은 연구 질문 및 포함된 특정 인구집단이나 세팅에 따라 다르다. 실용적인 제안은 다음과 같다.

- 예상되는 검색의 복잡성을 감안할 때 검색 전략을 계획하고 실행할 때 전문가 조언을 받도록 하라. 전문가는 출판되지 않았거나 색인되지 않았거나 찾기 어려운 근거를 알고 있을 수 있다.
- 민감도 및 특이도를 고려하여 검증된 필터를 찾아 알려진 제한점을 교정하려고 시도하라. 필터가 검증되지 않았다면 중요한 정보를 누락시킬 수 있는 위험이 있음을 신중히 고려하라.
- 전통적인 데이터베이스 외의 것을 찾도록 하라: 연구 주제를 다루는 규모가 작고 특징적인 데이터베이스가 더욱 적절할 수 있다.
- 검색 전략에 대한 결정을 명확하게 이해하기 위해 로직 모델을 개발하라.
- 반복적으로 검색하라.

16.4 근거 평가

형평성 연구의 경우 PROGRESS-plus 요인 사이의 기저상태 불균형(baseline imbalance)을 평가하면 질 낮은 무작위배정에 대한 체크를 할 수 있다. 또한 형평성 요소들이 비무작위 연구의 잠재적인 교란요인으로 고려될 수 있다. 저자는 PROGRESS-plus에서 정의한 특정 인구집단에서 탈락이 차별적으로 발생했는지 여부를 문서화해야 한다. 그 외에 근거의 비판적 평가는 다른 문헌고찰과 유사하다.

16.5 근거 합성(synthesis)

형평성 분석은 세 단계로 이루어진다: 첫째, 건강 불평등을 경험할 가능성이 있는 인구집단을 프로토콜에서 확인한다. 둘째, 중재가 중요한 개선을 가져올지 평가한다. 셋째, 확인된 인구집단이 다른 인구집단과 마찬가지로 절대적 효과와 상대적 효과 모두에서 동일한 개선을 보이는지 평가한다. 효과의 기울기(gradients of effects)와 절대 및 상대 효과의 차이(gap)을 평가하는 방법은 Evans et al(2007)에 기술되어 있다.

16.5.1 하위군 분석

어떤 중재가 광범위한 인구집단에 이루어지는 경우 형평성 연구에서 로직 모델 및 프로토콜에 사전에 정의된 PROGRESS-plus 요인 중 하나 이상의 하위군 분석을 고려할 수 있다.

모든 하위군 분석은 미리 정하고 정당화되어야 한다. 문헌고찰을 수행하는 과정에서 연구결과에 영향을 미치는 다른 중요한 요소가 밝혀질 수 있다. 저자는 이를 공개해야 하고 모든 사후 결정은 문서화되어야 한다.

메타 회귀분석를 통해 인구집단, 맥락, 과정 요인과 같은 설명변수의 역할을 평가할 수 있다.

16.5.2 '결과 요약표'

저자는 '결과 요약표'에 건강 불평등에 관한 연구결과를 통합하는 5가지 방법 중 하나를 고려할 수 있다.

1. 중재결과로서 건강 형평성을 포함한다.
2. 건강 형평성과 관련된 환자에게 중요한 중재결과(patient-important outcomes)를 고려한다.
3. 상대적 효과성의 중요한 차이를 강조하기 위해 건강 불평등을 경험한 인구집단을 위한 별도의 표를 제시한다.
4. 특정 인구집단에 대한 기저상태 위험의 차이를 강조하기 위해 단일 표 내에 다른 행을 작성한다.
5. 건강 불평등을 경험하는 사람들로 미리 정의된 인구집단에 대한 근거의 간접성(indirectness)를 평가한다.

16.6 근거 해석

PROGRESS-plus에 의해 정의된 특정 집단에 대한 근거 해석은 프로토콜 단계에서 중재를 받는 사람으로 확인된 인구집단에 초점을 맞추어야 한다. 해석할 때 다음과 같은 질문을 고려해야 한다: 포함된 인구집단에서 참여한 인구집단의 상당 부분을 차지하지 않더라도 해당 집단에 연구결과가 적용될 가능성이 있나? 그 이유는 무엇인가?

16.7 맺음말

문헌고찰 저자는 형평성 체크리스트(Ueffing et al 2009)와 같은 도구를 사용하여 제목과 프로토콜 단계에서 건강 형평성의 적절성을 명시적으로 고려한 다음, 건강 형평성에 대한 영향을 평가하고 일반화 가능성과 적용가능성을 토론하기 위한 방법을 고안할 것을 권고한다.

16.8 참고문헌과 읽을거리

- EPOC. LMIC Databases: Effective Practice and Organisation of Care; 2013. http://epoc.cochrane.org/lmic-databases.
- Evans T, Whitehead M, Diderichsen F, Bhuya A, Wirth M. Challenging Inequities in Health: from Ethics to Action. New York(NY): Oxford University Press; 2001.
- Kristjansson E, Francis DK, Liberato S, Benkhalti Jandu M, Welch V, Batal M, Greenhalgh T, Rader T, Noonan E, Shea B, Janzen L, Wells GA, Petticrew M. Food supplementation for improving the physical and psychosocial health of socio-economically disadvantaged children aged three months to five years. Cochrane Database of Systematic Reviews 2015; 3: CD009924.
- Marmot M, Friel S, Bell R, Houweling T, Taylor S, Commission on Social Determinants of Health. Closing the gap in a generation: health equity through action on the social determinants of health. Lancet 2008; 372: 1661-1669.
- O'Neill J, Tabish H, Welch V, Petticrew M, Pottie K, Clarke M, Evans T, Pardo Pardo J, Waters E, White H, Tugwell P. Applying an Equity Lens to interventions: Using PROGRESS to ensure consideration of socially stratifying factors to illuminate inequities in health. Journal of Clinical Epidemiology 2013; 67: 56-64.
- Pederson A, Armstrong P. Sex, gender and systematic reviews: the example of wait times for hip and knee replacements. In: Armstrong P, Pederson A, editors. Women's Health: Intersections of Policy, Research and Practice. Toronto: Women's Press; 2015. p. 56-72.
- Pfadenhauer LM, Gerhardus A, Mozygemba K, Lysdahl KB, Booth A, Hofmann B, Wahlster P, Polus S, Burns J, Brereton L, Rehfuess E. Making sense of complexity in context and implementation: the Context and Implementation of Complex Interventions(CICI) framework. Implementation Science 2017; 12: 21.
- Sinha IP, Altman DG, Beresford MW, Boers M, Clarke M, Craig J, Alberighi OD, Fernandes RM, Hartling L, Johnston BC, Lux A, Plint A, Tugwell P, Turner M, van der Lee JH, Offringa M, Williamson PR, Smyth RL. Standard 5: selection, measurement, and reporting of outcomes in clinical trials in children. Pediatrics 2012; 129 Suppl 3: S146-152.
- Welch V, Petticrew M, Tugwell P, Moher D, O'Neill J, Waters E, White H. PRISMA-Equity 2012 extension: reporting guidelines for systematic reviews with a focus on health equity. PLoS Medicine 2012; 9: e1001333.

제17장 중재의 복합성

핵심 포인트

- 이 장에서는 '복합중재'가 아닌 '중재의 복합성'을 언급한다. 왜냐하면 어떤 중재도 간단하지 않으며, 많은 체계적 문헌고찰 저자들은 복합성의 일부 측면을 고려해야 하기 때문이다.
- 중재의 복합성을 이해하는 세 가지 방법이 있다.

  (i) 중재에 포함된 구성 요소의 수

  (ii) 중재 요소 간의 상호작용 및/또는 중재와 그 맥락 간의 상호작용의 관점

  (iii) 중재가 도입된 더 넓은 시스템의 관점

17.1 개요

이 장에서는 체계적 문헌고찰 내에서 중재의 복합성을 개념화하고 고려하는 방법을 소개한다. 중재의 복합성을 고려하는 것이 체계적 문헌고찰 자체를 지나치게 복잡하게 하고 사용자가 이해하기 어렵게 할 수 있다는 위험이 있다. 이 장에서는 그런 상황 속에서 체계적 문헌고찰을 지원하는 접근법에 대한 이해와 언제 어떤 상황에서 이를 적용할지에 대한 결정을 내리는 데 필요한 정보를 제공하는 것을 목표로 한다. 이 장이 복합성의 모든 측면을 다루는 것을 목표로 하지 않지만, 저자들에게 중재의 복합성 문제를 해결하기 위한 문헌고찰의 질문을 만들고 추가 출처를 찾아보는 방법에 대해 설명한다.

17.1.1 중재의 복합성의 개념

'단순중재(simple intervention)', '복합중재(complex intervention)'라는 용어는 중재의 복합성을 다루는 많은 문헌에서 흔히 사용된다. '단순성'과 '복합성'은 단지 중재의 물리적 특성에 따라 단순과 복합이란 이분법적 범주로 구분하는 것이 아니기 때문에 구체적으로 중재의 복합성을 언급할 것이다. 흔히 단순하다고 생각하는 약물은 더 넓은 맥락에서 분석하는 경우(예: 환자-의사 관계 또는 약이 제공되는 보건의료체계 하에서)에서는 '복합중재'로도 볼 수 있다. 체내 약물의 약력학, 약동학 등을 고려한다면 약물의 복용이라는 단순한 중재조차도 복잡해진다. 복합성을 다차원 연속체로서 고려할 때, 중재의 서로 다른 측면에 걸쳐 더 높거나 더 낮은 수준의 복합성이 있을 수 있으므로, 이에 따라 문헌고찰의 저자가 복합성의 어떤 측면에

초점을 맞추는 것이 가장 중요한지를 결정하는 데 도움이 될 수 있다.

중재의 복합성에 대해 생각할 수 있는 세 가지 방법이 있는데, 이것은 중재와 맥락에 대한 대안적인 관점을 제공한다. 첫 번째와 두 가지 관점은 (i) 중재 자체가 어떻게 복잡할 수 있는지에 대한 문제와 (ii) 특정 상황에서 중재가 어떻게 복합적인 상호작용하는지에 초점을 맞추고 있다. (iii) 세 번째 관점은 개별적 중재보다 더 넓은 맥락의 실행관점이 초점이다.

중재의 복합성에 대한 첫 번째 개념은 가장 간단한 것으로 둘 이상의 구성요소를 가진 중재로 정의한다. 중재 중 어떤 구성요소가 가장 중요하고 실제 효과를 유발하는 요소인지를 이해하기 어려울 수 있기 때문이다. 이 경우 분석 방법은 각각의 구성요소가 부가적인(additive) 효과로 작용한다는 관점에 기초한다.

중재의 복합성에 대한 두 번째 관점은 중재의 구성요소 사이에서, 또는 중재와 연구대상자들 사이의 상호작용 혹은 둘 다에 초점을 맞춘다. 이러한 복합성을 이해하는 것은 두 가지 중요한 의미를 가진다, 즉 (1) 더 복합적 상호작용이 있으면 다른 분석 방법을 사용한다는 점(예를 들어, 한 요소의 용량이나 강도가 다른 요소가 활성화되기 전에 주어진 임계값에 도달해야 하는 경우), (2) 상당히 '단순'해 보이는 중재(예: 단일 약물의 처방)라도 환자가 보이는 치료순응도 같은 문제를 고려할 때 복합성은 발생한다는 점이 그것이다. 세 번째 관점에서 분석은 특정 중재와 중재결과에 대한 고려에서 중재가 도입된 더 넓은 맥락('시스템'으로 이해됨)으로 초점을 전환할 수 있다. 여기서 분석은 중재가 시스템에 미치는 영향 또는 시스템이 중재에 미치는 영향을 조사할 수 있다.

17.1.2 관점 1 및 2: 중재 내외부의 여러 요소 및/또는 상호 작용에서 발생하는 중재의 복합성

체계적 문헌고찰은 종종 복합성의 함의를 충분히 고려하지 않고 중재의 효과와 그 요소들의 결합에 대해 부가적 관점의 접근법을 채택한다. 그러나 중재의 각 구성요소는 종종(부가적 효과 대신) 상승(synergistic) 효과를 가질 수 있으며, 이는 중재의 복합성의 한 가지 주요 특징이다.

복합중재를 정의하고, 개발하고, 문서화하며 재현하는 데에는 특정한 어려움이 있다. 대표적인 예로 전문 뇌졸중 집중치료실의 장점을 평가하기 위한 임상시험 설계가 있다. 그러한 임상시험은 진단, 약물투여, 치료지침, 퇴원 및 후속 조치 등에 관여하는 다양한 전문가들의 전문성을 고려해야 한다. 또한 뇌졸중 집중치료실은 조직, 관리, 여러 관련 기술의 측면에서도 다를 수 있다. 뇌졸중 집중치료실의 다양한 구성 요소 중 특정 활성요소를 규정하기 어려울 수 있으므로 똑같은 중재를 다른 기관에 적용하기 어려울 수 있다. 중재의 복합성을 주요 측면을 설명할 수 있는 부가적인 세부사항의 예는 다음과 같다.

- 중재 내에 복수의 구성요소가 있는지 여부, 그리고 그것들이 서로 상호작용할 수 있는지 여부

- 중재를 제공하거나 받는 사람에게 요구되는 행동 범위 및 난이도와 변이성
- 중재가 비선형적인 효과를 보이는지 여부
- 중재 대상군의 수
- 중재결과 개수와 변동성
- 중재의 유연성 또는 조정가능성

17.1.2.1 맥락, 실행, 기전

'맥락'은 보통 복합성을 다루는 문헌에서 핵심 개념으로 설명되지만 별도로 정의하기는 어렵다. 종종 중재가 어떻게 실행되는지와 중재가 어떻게 작동하는지와 같은 관련 문제와 결합되어 있다. 사전적으로는 '맥락'을 다음과 같이 정의한다.

"사건, 상태 또는 아이디어에 대한 배경을 형성하고, 그것이 완전히 이해될 수 있는 상황"

이러한 용어로 정의될 때, 중재의 맥락을 알고, 그것이 어떻게 해당 중재결과를 초래했는지를 '완전히 이해'하는 것은 매우 도전적이다.

더욱이 '맥락'을 정의하는 것 자체가 판단의 문제라는 것도 도전적이다. 비무작위 연구의 평가를 위한 ROBINS-I 도구는 "보건의료 세팅"(예: 외래 환자 대 병원 환자), 적용 서비스(예: 관리의료 대 급여 프로그램), 지리적 환경(예: 시골 대 도시), 문화적 환경, 법적 환경" 등을 맥락적 특징으로 정의한다. 이 같은 맥락적 고려를 더욱 확대하여 중재 사이의 상호작용, 연구대상자 사이의 상호작용도 맥락으로 인정된다.

맥락은 능동적이고 고유한 요인으로 구성되는 일련의 특성 및 상황을 반영하며, 그 안에 구현이 포함된다. 이와 같이 맥락은 단지 실행의 배경이 아니라 중재의 실행과 능동적으로 상호작용하여 중재를 수정하거나 촉진하거나 제약한다.

따라야 할 특정한 일련의 절차에 따라 계획된 중재를 **실행**할 필요가 있다. **실행**은 다음과 같이 정의한다.

"특정 세팅 하에서 정책과 실천에 특정 중재를 가할 목적으로 적극적으로 계획되고 의도적으로 시작된 노력이다. 이러한 조치는 중재의 적극적 사용을 채택한 사람에 의해 이루어지며, 실행 이론에 의해 뒷받침 된다"

고려해야 할 중요한 측면에는 실행의 복합성(즉, 특정 연구에서 실행 프로세스의 변화에 의해 중재의 영향이 바뀔 것으로 예상되는 상황)과 연구대상자 반응의 복합성(즉, 특정 연구에서 연구대상자의 변이에 의해 중재의 영향이 바뀔 것으로 예상되는 상황)이 포함된다. 이런 변화하에서 어떤 실행은 효과가 있을 것이고 어떤 실행은 그렇지 않을 것이다. 체계적 문헌고찰 내

에서 이러한 서로 다른 중재 변화와 이의 실행을 비교할 수도 있을 것이다. 이러한 상황을 이해하기 위해서는 변이의 내용을 알 수 있도록 중재의 내용을 풀어서 살필 필요가 있을 것이다.

대부분의 경우 중재의 충실도(integrity)는 각 연구마다 중재를 최적 수준의 용량과 동일한 방식으로 제공하는 것으로 정의된다. 그러나 복합중재에서는 중재의 충실도는 다른 의미를 가지며 과정과 기능을 표준화하면서 형식이 수정될 수 있도록 하는 것이다.

중재의 형식과 기능을 구분하는 것의 의미는 우울증을 예방하기 위한 지역사회 교육 중재에 대한 군집 무작위배정 임상시험의 예로 설명할 수 있다. 군집간 중재의 '형식'을 유지하기 위해, 평가자는 모든 연구대상자에게 동일한 서면 정보가 제공되고 있는지 확인해야 한다. 반면 군집 전체에서 '기능'이 일관되도록 하기 위해, "각 군집의 읽고 쓰는 능력, 언어, 문화 및 학습 스타일"에 맞춘 중재를 전달하는 방법을 고안해야 하며 '기능'(또는 메커니즘)에 대한 충실도를 확보하기 위해, 원래의 중재 부분의 '형식'을 조정할 필요가 있다.

맥락에 의한 변화는 중재를 받는 연구대상자의 유형과 특성에 의해 영향을 받을 수 있어 '맥락'과 '세팅'을 분리하는 것도 어려울 수 있다.

중재의 예상 **작동기전**을 이해하고 설명하려면 중재의 이론적 근거를 이해하는 것이 좋다. 어떤 상황에서는 중재와 그 중재결과 사이에 비교적 잘 이해되는 인과 경로가 있는 경우가 많다. 예를 들어 특정 의료 중재와 중재결과의 사이의 생리적 경로는 기초 의학적 지식을 통해 이해할 수 있다. 그러나 다른 더 복잡한 상황(예: 중재가 맥락 안에서 상호 작용하고 변화에 적응하는 상황)의 경우, 그러한 경로는 잘 이해되지 않고, 예측성이 떨어지며, 비선형적일 수 있다. 문헌고찰을 시작할 때 중재가 어떻게 어떤 기전을 통해 작동할지에 대해 이론적 근거를 세우면 초기 가정을 명확히 하는 데 도움이 될 수 있다. 체계적 문헌고찰의 결과는 그 이론의 타당성을 검증할 뿐만 아니라 중재 이론을 발전시킬 것이다.

중재가 어떻게 작용하는지를 이해하려면 개별 구성요소를 식별하고 이러한 구성요소가 개별적으로 또는 조합하여 어떻게 그 효과를 발휘하는지 이해해야 한다. 또한 실행 맥락과 중재 실행에 관련된 프로세스에 대한 추가 검토가 필요할 것이다. 그 의미는 실행 과정의 변화에 의해 중재의 영향이 수정될 것으로 예상하는 상황은 문헌고찰대상 연구마다 다를 수 있다는 것이다. 또한, 중재를 받는 연구대상자 사이의 변이에 의해 중재의 영향이 바뀌는 상황도 연구마다 다를 수 있다.

고찰의 예

아동과 청소년의 천식에 대한 학교 기반 자기 관리 중재의 코크란 리뷰는 중재 효과와 중재가 어떻게 실행되었는지, 그리고 다른 군간 관찰된 영향의 차이를 어떻게 설명할 수 있는지를 이해하는 데 관심이 있었다. 천식에 대한 교육의 영향에는 사회경제적 격차가 있었는데, 사회경제적 하위집단 및 소수민족 아이들이 천식과 관련된 입원이 더 많았다. 그 이유 중 하나는 학교 기반 자기 관리 중재의 차등 효과일 수 있다. 사회경제적 불평등이 환경에 뚜렷이 나타난다는 점을 감안할 때, 이러한 문제들은 순전히 개별 연구대상자의 특징으로 이해될 수 없다.

17.1.3 관점 3: 복잡한 시스템 내의 중재

시스템 관점은 중재가 분리된 사건이나 구성요소의 집합이 아니라 상호 연결된 시스템의 일부 또는 '사건'으로 본다. 따라서 문헌고찰은 중재가 기존 시스템과 상호 작용하는 광범위한 맥락 안에서 이해하는 것을 목표로 하며, 시스템이 어떻게 변경되는지 또는 시스템이 중재에 어떻게 영향을 미치는지 또는 둘 다에 대해 검토한다. 이러한 관점을 이용하여 문헌고찰을 실시할 때에는 중재의 구성요소를 고려할 뿐만 아니라, 중재가 도입되는 시스템을 규정할 필요가 있을 것이다. 예를 들어, 저소득 또는 중간 소득 국가에서 적시에 새로운 백신을 도입하는 것은 백신의 공급(정부 및 공급사 사이의 합의 포함), 냉장상태 유지를 위한 의료 시설의 업그레이드(예: 냉장고 공급), 보건 종사자 교육, 기존의 통상적인 예방접종 시스템(예: 표준 소아 예방접종)과는 별개로 백신을 제공하는 시스템 등을 포함한다. 이는 기존의 의료 시스템 하에서의 다른 중재에 의도하지 않은 방식으로 영향을 미침으로써 시스템 전체에 긍정적 또는 부정적 영향을 미칠 수 있다. 시스템적 관점을 취하는 문헌고찰은 시스템의 기능과 시간이 지남에 따라 어떻게 변화하는지, 그리고 그러한 시스템 변화에 대한 중재의 기여에 대한 광범위한 질문에 대한 해답을 목표로 할 수 있다. 완전한 설명은 이 장의 범위를 벗어나나 체계적 문헌고찰에서 복잡한 시스템 관점의 역할은 여전히 진화하고 있다.

저자는 체계적 문헌고찰 안에 시스템적 관점에 따른 내용을 추가할지를 결정할 때 다음과 같은 질문을 고려해야 한다.

- 문헌고찰 사용자가 알고자 하는 내용은 무엇인가? 중재인가 시스템인가? 아니면 둘 다인가?
- 어느 수준에서 중재가 이루어졌는가? 중재의 효과가 개인을 넘어서는 수준에서 영향을 미칠 가능성이 있는가? 만일 실행과 효과가 가족, 지역사회 또는 그 이상으로 파급되는 경우라면 시스템적 관점을 취하는 것이 도움이 될 수 있다.
- 중재가 (i) 개별적이고 식별 가능한 중재 또는 중재 패키지인가, 아니면 (ii) 기존 시스템 내에서 더 폭 넓은 변화인가?

문헌고찰 저자는 또한 문헌고찰을 수행하는 데 사용할 수 있는 가용자원을 고려해야 한다. 더 넓은 시스템 내에서 중재에 대한 대규모 문헌고찰은 시간이 많이 걸리고 비용이 많이 들며 다수의 다학제 팀이 필요할 수 있다. 그것은 또한 사용자의 요구수준을 넘어서는 복잡한 대답을 만들어 낼 수도 있다.

17.1.4 주요 사항 요약

중재의 복합성을 이해하는 세 가지 방법이 있다: (i) 중재의 구성요소 수 측면에서, (ii) 중재의 요소 및/또는 중재와 그 맥락 사이의 상호작용; 그리고 (iii) 중재가 도입된 더 넓은 시스템의 관점에서.

중재의 복합성을 고려할 때, 중재의 작동기전, 그것이 도입된 맥락 및 실행과 관련된 문제에 특히 주의를 기울일 필요가 있다. 문헌고찰 팀은 아래 중 복합성에 대한 어떤 관점이 문헌고찰과 관련이 있는지 검토해야 한다.

- 문헌고찰이 여러 구성 요소로 구성되는 중재를 다루고 있는가?
- 관심 중재가 실행되는 맥락과 상호 작용하고 중재의 적응(adaptation)이 이루어질 가능성이 있는가?
- 어떤 분석 방법을 사용해야 하는가(예: 상호작용 및/또는 비선형 효과 모델링에 적합한 방법)?
- 행동, 맥락, 실행의 핵심 개념은 어떻게 정의되고 있는가?

17.2 문헌고찰의 구조화(formulation)

복합성을 해소하기 위해서는 다음과 같이 효과 외적인 문제에 대한 질문을 해야 한다.

- 어떤 상황에서 중재가 효과가 있는가?
- 다중 요소 중재에서 각 요소들 간의 상대적 중요성과 시너지 효과는 무엇인가?
- 중재가 효과를 달성하는 작용 기전은 무엇인가?
- 실행 및 연구대상자의 반응에 영향을 미치는 요인은 무엇인가?
- 다른 맥락에서 중재의 타당성과 수용가능성은 어떻게 바뀌나?
- 시스템적 관점 하에서의 역학 관계는 어떠한가?

그러므로 체계적 문헌고찰은 중재(예를 들어, 중재의 구성요소가 상호작용하는 방법)의 관점에서 복합성을 고려할 수 있으며, 또한 그것이 구현되는 방식에서도 복합성을 고려할 수 있다. 이 상황에서 체계적 문헌고찰은 복합성의 개념을 사용하여 실행 이론의 개발에 도움을 줄 수 있으며, 실행을 개선하기 위한 전략을 알려줄 수 있다.

광범위한 문헌고찰 질문을 다루는 것은 검색 전략, 근거 유형, 적격성 기준, 근거 평가, 문헌고찰 설계, 합성 방법에 영향을 미친다. 때때로 관심 질문을 다루기 위해 둘 이상의 유형의 연구설계가 필요할 수 있으며, 연구결과물들은 제 3의 합성물로 통합될 수도 있다.

17.2.1 이론 및 로직 모델의 역할

저자는 관심 중재가 어떻게 작용할 것으로 예상하는지를 프로토콜에 명시해야 한다. 인과적 경로가 잘 이해되는 문헌고찰이라면 이것은 적절한 문헌을 단순히 참조하는 비교적 간단한 과정이 될 수 있다. 중재 사이에 복합성 또는 다양성이 많은 문헌고찰에서 로직모델 중재의 잠재적 작용기전을 설명하는 인과 경로 및 중재자(mediator)와 조정자(moderator)를 도식으로 표현하는 데 사용된다.

17.2.2 중재의 복합성을 해결하기 위한 질문의 공식화

문헌고찰을 진행할 때는 목적이 매우 중요하며 그런 의미에서 단순히 '이 중재가 효과가 있는가?'라고 묻는 것보다, 중재의 어떤 방식이 다른 방식에 비해 더 적합할 수 있는 상황인가와 같은 방식으로 질문으로 만드는 것이 더 유용하다. 앞에서 언급한바와 같이 문헌고찰 질문 작성을 위해서는 문제를 개략적으로 제시하고, 문헌고찰 대상이 되는 중재의 맥락적 요인의 우선순위를 정하는 것이 중요하다. 중재의 복합성이 큰 상황에서는 독자 및 다른 이해당사자가 문헌고찰에서 필요한 상황적 요인을 식별하는 데 어떻게 도움을 줄 수 있는지를 고려해야 한다. 각 그룹의 사람들마다 상당히 다양한 요구를 할 수 있고 그것들을 한 번의 문헌고찰로 모두 고려하는 것은 불가능할 수도 있다. 질문의 우선순위를 정하기 위한 공식적인 방법론을 포함하여 이해당사자들을 참여시킬 수도 있다. 때때로 문헌에서 중재가 정의되는 방법을 완전히 이해하기 위해 공식적인 scoping review를 수행할 수도 있다.

중재의 특정 측면이나 실행, 넓은 맥락이 중요한가를 고려할 때, 문헌고찰 저자는 특정 변이성은 종종 불가피하며 상상할 수 있는 모든 차이를 조사하는 것은 불가능할 것이라는 것을 기억해야 한다. 특히, 중재의 복합성의 모든 측면이 문헌고찰 질문에서 상세히 설명되어야 하는 것은 아니다. 로직 모델과 합성을 위해 확인된 모든 하위군 내에서 이러한 측면을 고려하는 것으로 충분할 수 있다.

17.2.3 PICO 및 복합성

PICO 프레임워크는 연구 질문의 틀을 만드는데 널리 사용된다. PICO 요소는 중재의 복합성이 예상될 때는 더 복잡해질 수 있다.

문헌고찰에서 고려되는 인구집단은 일반적으로 건강 상태(예: 골다공증 환자) 또는 행동(예: 흡연하는 청소년)의 측면과 관련 인구통계학적 요인과 세팅으로 설명된다. 인구집단의 변화에 초점을 맞춘 연구에서는 인구집단의 정의 자체가 혼란스러울 수 있다. 결정적으로, 인구집단은 개인의 특징만을 종합한 것이 아니라, 사회적(그리고 물리적) 관계가 인구의 건강 분포를 형성하기도 한다(사회망을 통해 비만이 전파되는 것이 좋은 예이다). 문헌고찰의 관심사는 인구집단 전체가 될 수도 있고, 인구집단 내의 서로 다른 군에 따라 영향이 어떻게 달라지는 가에 대한 것일 수도 있다.

중재와 관련하여, 핵심 과제는 중재의 정의에 있다. 중재의 복합성을 고려할 때, 문헌고찰 저자는 중재가 실행될 수 있는 광범위한 방법을 고려해야 하며, 기능이 유사하더라도 형태가 다르기 때문에 특정 중재가 일차 평가에서 배제되는 것을 피해야 한다.

문헌고찰에서 비교 중재는 세심한 고려가 필요할 수 있다. 특히 과세, 규제 또는 환경 변화와 같은 구조적 중재가 평가되고 있는 경우 또는 각 중재 대상군이 복잡한 경우 적합한 대조군을 파악하는 것이 어려울 수 있다. 문헌고찰 저자는 특히 시스템 효과로 인해 발생할 수 있는

교란 요인에 유의해야 하며, 여기서 더 넓은 맥락적 요인이 특정 상황에서 중재의 영향을 감소시키거나 향상시킬 수 있다.

관심 중재결과에는 의도한 것과 의도하지 않은 것 그리고 건강에 관련된 것과 관련이 없는 것으로 구분할 수 있다. 우선순위에 따라 중재결과를 선택하는 것은 판단과 관점의 문제로 선택 결정에 대한 근거를 명시적으로 보고해야 한다. 문헌고찰 저자는 중재결과의 우선순위가 문화적으로 다양하며, 다양한 사람들의 관점 즉 연구대상자(예: 환자, 고위험군), 중재 제공자(예: 임상의사, 병원 직원 등), 또는 정책결정자, 중재 결정자, 재정지원자, 일반 대중의 각각의 관점에 따라 달라질 수 있음을 주목해야 한다.

17.2.4 맥락 및 실행을 다루기

중재 복합성의 한 가지 중요한 측면은 중재 효과가 종종 맥락에 크게 의존하며, 맥락은 효과의 조정자(moderator, 즉, 강도나 방향에 영향을 주는 것)와 효과의 중재자(mediator, 즉, 효과가 왜 관찰되는 가를 설명하는 것)로 작용한다는 것이다. 이는 GRADE 평가를 적용할 때 문헌고찰 결과의 적용성(applicability)를 판단하는 데 영향을 미친다. 가장 일반적인 어려움 중 하나는 중재가 서로 다른 맥락에서 서로 다른 영향을 미친다는 것이며, 따라서 문헌고찰 저자는 문헌고찰의 초점을 하나의 특정 맥락이나 세팅(예: 학교에서 수행된 연구)으로 제한하는 것이 더 의미 있는 것인지에 대해(고찰의 이해당사자와 협의하여) 견해를 가질 필요가 있을 것이다. 일부 문헌고찰의 경우, 다른 맥락에서 중재와 중재의 영향을 이해하는 것이 문헌고찰을 수행하는 주요 이유일 수 있으며, 그러한 경우 문헌고찰 저자는 문헌고찰을 계획하고 수행할 때 맥락을 고려할 필요가 있을 것이다. 고려해야 할 맥락에는 (i) 문헌고찰 질문의 맥락, (ii) 포함된 연구의 맥락, (iii) 문헌고찰 결과 또는 권고사항이 도입될 구현 맥락 등이 다양하게 존재한다. 특정 상황에서 의사결정에 정보를 제공하고자 하는 경우, 근거의 '직접성(연구대상자, 중재 및 중재결과가 유사한 정도)에 대해 고려해야 한다.

17.2.5 어떤 유형의 연구들이 중재의 복합성을 다루는가?

어떤 연구설계를 포함할지에 대한 결정은 문헌고찰의 질문에 대한 답변을 위한 연구의 '목적 적합성(fitness for purpose)'을 고려해야 한다. 대부분은 중재와 중재결과 사이의 인과 관계를 규명하는 연구설계상의 강점 때문에 무작위배정 임상시험의 연구결과를 종합하는 데 초점을 맞춘다. 그러나 모든 유형의 중재를 대상으로 무작위배정 임상시험이 항상 가능한 것은 아니므로, 중재 효과와 중재 요소들 간의 상호작용에 대한 근거도 무작위배정 임상시험, 준실험연구, 비무작위연구 등에서 도출될 수 있다. 비무작위로 설계된 대규모 연구 혹은 정책 기반 중재(예: 지역 기반 재생 프로그램)는 비교가능한 대조군을 이용할 수 없거나 별도의 대조군이 없는 전후연구, 시계열연구, 준실험연구 같은 연구 디자인만 가능할 수 있다. 비무작위

연구와 대조군 없는 연구를 제외하는 것은 존재하는 연구를 평가에서 배제하는 결과로 이어질 수 있으며, 어떤 경우에는 그러한 설계가 적절한 효과의 근거를 제공할 수 있다. 예를 들어, 흡연 금지가 관상동맥 질환 입원에 미치는 영향을 평가할 때, 단절적 시계열 연구를 채택할 수 있다.

복합성을 다루는 체계적 문헌고찰에서 질문은 중재의 효과 여부를 묻는 것을 넘어서 그것이 어떻게 작동하는지, 어떤 상황에서 어떤 사람에게 효과적인지 묻는 경우가 많다. 이러한 문제를 해결하기 위해서는 다양한 연구설계가 포함될 수 있다. 특히, 중재가 중간 및 최종 중재결과에 영향을 미치는 과정에 대한 근거뿐만 아니라 중재의 수용성 및 실행에 대한 질적 근거가 포함되는 경우가 많다. 질적 근거는 주요 정량적 평가 연구에서는 보고되지 않을 수 있는 의도하지 않은 유해 근거를 확인할 수 있다.

17.3 근거의 확인

중재의 복합성을 탐구하는 데 초점을 맞춘 문헌고찰에 포함할 근거를 검색하는 것에 대한 상세한 지침은 거의 없다. 핵심이라고 할 수 있는 것은 그러한 문헌고찰을 위해서는 정성적 및 정량적 근거를 포함한 다양한 질문에 답할 수 있도록 해야 한다는 것이다. 중재가 실행되고 달성되는 것과 관련이 있는 이론, 맥락, 프로세스 및 기전에 대한 정보 검색도 필요할 수 있다.

이를 위해서는 그러한 자료가 어디에 있는지(예를 들어, 표준적인 문헌 자료원 이외의 출처 포함), 가능한 연구설계, 문헌고찰에 대한 검색 및 방법론적 의사결정을 안내하는 이론의 역할에 대한 검토가 필요하다. 정책 문서, 정성적 자료, 표준적인 문헌 자료원 외부에 있는 출처 및 관련 지식을 갖춘 자문 그룹과의 논의도 유용한 정보를 제공할 수 있다. 실제로, 연구 활동의 범위를 이해하기 위해 광범위한 검색을 수행하는 체계적 문헌고찰 내에 별도의 '지도(mapping)' 단계를 포함하는 것은 문헌고찰 프로세스에 매우 귀중한 추가 단계가 될 수 있다. 이 근거에 대한 일부 사전 문헌 고찰을 통해 중재가 어떤 형태를 취하는지, 변화의 목표 수준이나 구조가 무엇인지, 그리고 그 목적이 무엇인지, 그리고 어떻게 변화를 가져올 것으로 기대되는지에 정보를 얻는데 도움을 받을 수 있다.

"군집 검색"(cluster searching)이라는 개념이 제안되었는데, "다양한 검색 기법을 사용하여 단일 연구와 관련된 논문이나 기타 연구 결과를 식별하려는 시스템적 시도"이다(Booth et al 2013). 이는 '핵심' 효과 연구와 직접 및 간접적으로 관련된 연구의 클러스터가 맥락, 수용성, 그리고 중재가 관심 결과에 영향을 미치는 프로세스를 알리기 위해 위치한다는 것을 의미한다. 이러한 문제에 대한 고려는 중재의 복합성을 이해하는 데 중요한 경우가 많으므로 문헌고찰 저자는 여러 출판물 사이에 분산될 수 있지만 포함된 연구에 대한 모든 관련 정보를 고려할 필요가 있다. 클러스터 검색 외에도, 포함된 중재 연구와 무관한 질적 및 프로세스 평가 연구를 더 폭넓게 검색하면, 문헌고찰 저자가 중재실행과 같은 더 넓은 측면을 다룰 수 있는 더 큰 근거의 풀을 만드는 데 도움이 될 수 있다.

17.4 근거 평가

중재의 복합성을 다루는 문헌고찰은 근거의 '목적 적합성' 개념에 초점을 맞출 필요가 있다. 즉, 연구 질문에 답하기 위해 어떤 유형의 증거가 가장 적합한지를 고려할 필요가 있다. 앞에서 기술한 바와 같이, 이러한 질문들은 중재의 실행 가능성 및 수용성에 관한 질문, 중재가 변화를 가져오는 프로세스와 기전에 관한 질문을 포함할 수 있다. 중재의 효과를 평가하는 연구를 평가할 때 무작위배정 임상시험에는 Cochrane ROB 2.0을, 비무작위 연구에는 ROBINS-I 도구를 사용해야 한다.

17.5 근거의 합성

초기에는 통계적 합성이 계획되어 있더라도, 일반적으로 비정량적 방법을 사용하여 합성을 시작하고, 문헌고찰에 포함된 인구집단과 중재의 특성을 이해하고, 이용 가능한 연구의 중재결과 자료를 구조적으로 검토하는 것이 유용하다. 이 점에서 표와 그림 도구는 저자가 복합성을 시각화하고 탐색하는 데 도움이 되는 중요한 역할을 할 수 있다. 여기에는 harvest plots, box-and-whisker plots, bubble plots, network diagrams, forest plots 등이 포함된다.

표준 메타 분석 방법은 중재군과 대조군의 합리적인 비교가능성에 의존하기 때문에 이질성이 높은 근거를 합성할 때는 적절하지 않을 수 있다. 그러나 중재와 인구집단이 질문에 답할 수 있을 정도로 유사한 경우, 특히 작용 기전을 명확하게 이해할 때 표준 메타 분석, 메타 회귀분석 또는 네트워크 메타 분석과 같은 접근방식이 적절할 수 있다.

문헌고찰 질문이 중재가 어떻게 그 효과를 달성하는지를 고려해야 하는 경우, 무작위배정 임상시험 이외의 다른 유형의 근거는 중재와 결과 사이의 인과 관계에 대한 이론을 제공하는 데 필수적이다. 로직 모델은 분석에 포함할 요인을 선택하는 데 대한 몇 가지 근거를 제공할 수 있지만, 복합성을 적절히 설명하기 위해 질적 근거의 추가 합성이 필요할 수 있다. 이는 특히 저자가 중재를 받는 사람의 관점과 경험을 고려해야 하는 차등적 중재 효과를 이해하는 경우에 해당된다. 이 모델은 중재, 연구대상자 및 맥락 사이의 상호작용이 어떻게 중재결과를 산출할 수 있는지를 요약하는 것을 목표로 하고 있다. 원인 체인 분석(causal chain analysis)은 메타 분석 경로 분석(meta-analytic path analysis) 및 구조 방정식 모델링(structural equation modelling), 모델 기반 메타 분석(model-based meta-analysis)을 포함하여 다양한 방법을 사용하여 할 수 있다.

동일한 분석에서 서로 다른 유형의 자료를 통합하는 것은 도전적이지만 강력한 접근방식이 될 수 있으며, 정량적 연구 사이의 이질성을 탐구하고 설명하는 데 질적 문헌의 합성을 위해 생성된 이론들을 사용할 수 있다.

복수의 구성요소를 가진 문헌고찰과 분석은 종종 순차적으로 복합성과 관련된 다양한 질문을 다루며, 각 구성요소는 이전 구성요소의 중재결과에 대해 설명될 수 있다. 사용된 방법으

로는 mixed-methods synthesis(qualitative thematic synthesis, meta-analysis, cross-study synthesis 등을 포함), Bayesian synthesis(정성적 연구가 사전정보로 이용되는 경우), qualitative comparative analysis 등이 있다. 그러한 분석은 차등적 중재 효과를 식별하기 위한 설명적 분석이며, 그러한 결과가 발생하는 이유를 설명하기 위한 것이다.

17.6 근거의 해석

다른 체계적 문헌고찰과 마찬가지로, 복합성에 초점을 맞춘 문헌고찰도 의사결정자를 돕기 위한 것이다. 따라서 근거의 본질을 고려하여 중재결과를 명확히 기술해야 한다. 중재결과가 다른 맥락과 환경에서 어떻게 적용될 수 있는지에 대한 고려도 중요하며, 아마도 중재의 복합성을 해결할 때 더욱 그러할 것이다. 중재의 영향은 맥락에 따라 유의하게 변화될 수 있으며, 문헌고찰 저자는 의사결정자가 중재의 시행 여부와 시행 방법을 결정할 때 고려해야 하는 맥락의 핵심 측면이 무엇인지 설명할 수 있다. 이는 여러 시나리오를 기술하고 연구 전체에 걸친 중재결과에서 이질성의 이유를 명확하게 기술함으로써 문헌고찰에서 명시적으로 수행할 수 있다. 복합성에 초점을 맞춘 문헌고찰의 잠재적 위험은 모든 중재의 실행이 다르게 보일 수 있다는 것이다. 의사결정자는 다음과 같은 결론을 내리기가 쉽다. 흥미롭지만, 근거는 전혀 없다. 그러나, 다른 문헌고찰에 대해서는, 문헌고찰에서 나온 근거를 새로운 결정을 알리는 데 사용할 수 없는 설득력 있는 이유가 있는지 생각해 보는 것이 도움이 될 것이다. 요컨대, 복합성(중재와 실행 차원에서) 때문에 항상 상황적 차이가 있겠지만, 이것이 근거를 쓸모없게 만들지는 않는다. 오히려 저자는 이 근거(여러 다른 맥락에 걸친 중재의 영향)를 새로운 결정을 알리기 위해 어떻게 사용할 수 있는지 고려할 필요가 있다. 예를 들어, 문헌고찰은 효과 크기 추정치의 범위를 보여줄 수 있고, 또는 이전 연구의 설정에 따라 예상 및 예상되지 않은 중재결과의 유형이 어떻게 달라지는지를 보여줄 수 있으며, 따라서 의사결정자는 향후 실행에서 가능한 조정 요인뿐만 아니라 가능한 반응의 범위에 대한 아이디어를 제공할 수 있다.

17.6.1 보고 지침 및 체계적 문헌고찰

중재의 복합성을 고려하는 체계적 문헌고찰은 그 자체로 복잡하며, 다양한 방법을 사용하여 다양한 유형의 근거를 통합한다. 체계적 문헌고찰을 위한 PRISMA-CI로 알려진 PRISMA 보고 지침의 확장이 발표되었는데 이것은 주로 정량적 근거에 초점을 맞춘 중재를 기술하기 위한 TIDieR 체크리스트를 보완하고 있으며, '복합중재'의 방법과 결과 보고에 대한 구체적인 지침이다. 관련 확장 항목들은 '복합중재'을 포괄하는 것으로 명확하게 식별하고, 문헌고찰에서 고려 중인 복합성의 특정 요소에 대한 정당성을 제공하며, 중재 또는 그 맥락의 복합성의 측면을 설명하는 것과 관련이 있다. ENTREQ 및 eMERGe 보고 지침은 정성적 근거 합성과 meta-ethnography을 보고하기 위한 것이다. 혼합 방법 문헌고찰(mixed-method reviews)

의 경우, 현재 지침이 존재하지 않지만, 기존의 정량적 및 정성적 보고 지침에서 적절한 보고 기준을 통합하기 위한 접근법을 제안한다. 무작위배정 임상시험 이외의 다양한 연구설계를 통합하여 복합성을 해결할 때 문헌고찰 저자가 직면해야 하는 과제 중 하나는 근거에 대한 GRADE 평가가 일반적으로 '낮음'으로 판명될 수 있으며, 포함된 여러 연구에 대한 상대적 신뢰도를 이해하는 측면에서 독자들에게 거의 도움이 되지 않는다는 점이다. 체계적 문헌고찰에서 합성된 근거의 양과 범위를 증가시키면 보고서를 읽기가 상당히 어렵고 긴 시간이 걸릴 수 있다. 결론에 충분히 명확한 보고서를 작성하는 것은 많은 수정을 필요로 할 수 있으며, 문헌고찰에 참여하는 소비자와 기타 이해당사자로부터 피드백을 얻는 것도 유용하다. 중재의 복합성은 이 단계 문헌고찰 중 필요한 자원을 증가시킬 수 있으며, 문헌고찰의 보고가 의사결정에 정보를 제공하는 데 사용될 수 있도록 충분히 명확해야 하는 경우 이를 위한 계획을 세우는 것이 필수적이다.

17.7 참고문헌과 읽을거리

- Booth A, Harris J, Croot E, Springett J, Campbell F, Wilkins E. Towards a methodology for cluster searching to provide conceptual and contextual "richness" for systematic reviews of complex interventions: case study(CLUSTER). BMC Medical Research Methodology 2013; 13: 118.
- Caldwell D, Welton N. Approaches for synthesising complex mental health interventions in meta-analysis. Evidence-Based Mental Health 2016; 19: 16-21.
- Khuder SA, Milz S, Jordan T, Price J, Silvestri K, Butler P. The impact of a smoking ban on hospital admissions for coronary heart disease. Preventive Medicine 2007; 45: 3-8.
- Noyes J, Hendry M, Booth A, Chandler J, Lewin S, Glenton C, Garside R. Current use was established and Cochrane guidance on selection of social theories for systematic reviews of complex interventions was developed. Journal of Clinical Epidemiology 2016b; 75: 78-92.
- Thomas J, Petticrew M, Noyes J, Chandler J, Rehfuess E, Tugwell P, Welch VA. Chapter 17: Intervention complexity. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.0 (updated July 2019). Cochrane, 2019. Available from www.training.cochrane.org/handbook.

제18장 개별 연구대상자 자료 분석

핵심 포인트

- 개별 연구대상자 자료(IPD) 문헌고찰은 각 연구의 연구대상자에 대한 원본 자료를 수집하고 재분석하는 유형의 체계적 문헌고찰이다. 자료는 연구자 또는 자료 공유 저장소 또는 플랫폼을 통해 얻을 수 있다.
- 이용 가능한 간행물 또는 기타 집계 자료(aggregate data) 품질에 문제가 있거나 철저한 분석이라고 볼 수 없는 경우 IPD 문헌고찰을 고려해야 한다. 어떤 상황에서는 IPD 합성이 적절한 첫 번째 단계가 될 수 있다. IPD 접근방식은 이용 가능한 자료의 품질을 개선하고 개별 연구의 불충분한 보고를 상쇄할 수 있다. 비뚤림위험은 더욱 철저하게 평가할 수 있으며, IPD는 집계 자료의 체계적 문헌고찰보다 더 상세하고 유연한 분석이 가능해 진다. IPD에 대해 접근하면 자료를 분석하고 결과를 다양한 방법으로 보고할 수 있는 가능성을 제공하므로 분석 방법을 자세히 미리 명시해야 하며, 보고 방법은 PRISMA-IPD를 따라야 한다.
- IPD 문헌고찰은 대개 동일한 문제에 대한 기존의 체계적 문헌고찰보다 더 오래 걸리고 비용이 더 많이 들며, 자료를 얻고, 관리하며, 분석하기 위한 다양한 기술이 필요하다. 따라서 상당한 시간과 자금 지원 없이는 하기 어렵다.

18.1 소개

18.1.1 IPD 문헌고찰이란?

개별 연구대상자 자료(Individual participant data, IPD)를 포함하는 체계적 문헌고찰에는 각 적격 연구의 원본 자료가 포함된다. IPD에는 일반적으로 연령, 성별, 건강상태의 성격과 같은 각 연구대상자에 대한 익명화된 인구통계학적 정보와 받은 치료나 검사, 관찰된 중재결과에 대한 정보를 포함한다. 이러한 자료를 중앙집중식으로 확인하고 분석할 수 있으며, 적절한 경우 메타 분석으로 결합할 수 있다. 일반적으로 IPD는 연구자에서 얻지만, 자료 공유 플랫폼과 자료 저장소를 통한 접근이 향후 몇 년 내에 가능할 수 있다.

집계 자료(aggregate data)와 비교하여 IPD를 수집하면, 더 많은 임상시험, 더 많은 연구대상자 및 중재결과를 포함함으로써 자료의 양과 품질을 상당히 개선할 수 있다. IPD는 또한 치료, 노출 또는 진단검사의 효과에 연령 또는 질병 심각도와 같은 연구대상자 수준의 공변량이 어떻게 영향을 끼치는지 조사할 수도 있다. 그렇게 함으로써 분석 범위를 넓히고, 유연성을 향상시켜준다. IPD 문헌고찰은 이와 같이 더 나은 품질의 자료와 분석을 가능하게 해주어 심층 탐색 및 강력한 메타 분석 결과를 제공하는데 도움이 될 수 있으며, 이는 집계 자료에 기반한 결과와 다를 수 있다. IPD 문헌고찰은 임상 진료와 연구에 상당한 영향을 미쳤으며, 치료 지침과 신규 연구에 더 많이 적용될 수 있다. 그러나 IPD 문헌고찰은 다른 문헌고찰보다 더 오래 걸릴 수 있다. 치료적 중재의 영향을 평가하는 문헌고찰은 일반적으로 완료하는 데 최소 2년이 걸린다. 또한, 전담으로 참여하는 팀원과 상당한 자금후원이 필요하다.

이 장에서는 IPD 수집이 문헌고찰에 유용하고 실현 가능한지 여부를 저자가 결정하는 데 도움이 되는 IPD 접근방식의 체계적 문헌고찰에 대한 개요를 제공한다. 대부분의 IPD 문헌고찰이 중재의 유효성을 평가하고 무작위배정 임상시험에 기초해 왔기 때문에, 이장에서도 중재의 효과, 무작위배정 임상시험에 초점을 맞추어 검토한다. 그러나 이 접근방식은 진단 및 예측 연구의 합성에 대해서도 특별한 장점을 제공하며, 설명된 많은 원칙이 이러한 종류의 합성에 적용될 수 있을 것이다. 이 장은 다른 장에서 설명된 방법에 대한 자세한 지침을 제공하지 않는다. 따라서 코크란 리뷰의 일부로 첫 번째 IPD 메타 분석을 수행하는 경우, IPD 메타 분석 방법 그룹을 통해 적절한 조언과 지침을 구해야 한다.

18.1.2 IPD 문헌고찰과 표준 체계적 문헌고찰은 어떻게 다른가?

IPD 문헌고찰에 대한 일반적인 접근방식은 집계 자료 체계적 문헌고찰과 동일하며, 차이점은 자료 수집, 점검 및 분석과 관련이 있다. 따라서 상세한 프로토콜이 준비되어야 하며, 리뷰 목적, 다룰 핵심 질문, IPD의 적격기준, 어떤 중재결과 자료가 수집될 것인지, 그리고 이를 어떻게 관리할 것인지, 그리고 계획 중인 분석과 같은 것들을 포함해야 한다. IPD 문헌고찰은 더 많은 분석의 가능성을 제공하기 때문에, 원하는 결과를 얻을 때까지 반복적으로 자료를 조회할 위험이 더 크다. 따라서 분석 방법을 프로토콜의 일부 또는 별도의 분석 계획에 미리 명시하는 것이 특히 중요하다.

일차 연구자를 참여시키면 해당 연구자에 의해 수행되거나 알려진 적격 연구를 추가할 수 있으며, 포함된 연구의 설계와 수행을 명확히 함으로써 비뚤림위험 평가의 신뢰성을 향상시킬 수 있다.

프로젝트는 가능한 경우 PRISMA-IPD에 따라 구조화된 보고서를 작성하고 배포하는 것으로 마무리되어야 한다. PRISMA-IPD는 PRISMA를 IPD에 맞도록 확장시킨 것이며, 효과성 문헌고찰에 초점을 맞추고 있지만, 다른 유형의 IPD 문헌고찰에도 많은 요소가 적용된다.

IPD에 기초한 체계적 문헌고찰에는 자료 관리 및 통계 분석과 같은 전문지식과 연구 협력

관리 기술이 필요하며, 동일한 문제에 대해 시행된 기존의 집계 자료 체계적 문헌고찰보다 더 오래 걸리고 더 많은 자원을 필요로 하는 경우가 많다. 따라서 IPD 문헌고찰은 '예비 시간'에 실시하기 어려우며, 전담 인력이 필요할 가능성이 있다.

18.1.3 IPD 문헌고찰은 어떻게 조직되는가?

IPD 문헌고찰은 대개 연구 자료를 제공하는 연구자과 프로젝트를 관리하고 수행하는 연구팀의 협업 프로젝트로서 수행된다. 이 협업 그룹은 설계와 수행의 모든 측면을 주도하고 책임지는 프로젝트 팀, 임상 및 방법론적 지침을 제공하고 전략적 의사결정을 지원하는 자문 그룹, 그리고 시험 정보와 IPD를 제공하고 원고 초안에 대한 의견을 제시하는 임상시험 연구자로 구성될 수 있다. 연구자 또는 기득권을 가진 단일 그룹이나 회사가 주도하는 프로젝트는 비뚤림의 위험이 더 크며, 신뢰성이 떨어진다고 볼 수 있다. 연구결과는 보통 모든 협력자가 출판물의 공동저자로 나열되어 발표되며, 모든 기여 내용과 이해상충은 명확하게 설명되어야 한다.

18.1.4 IPD 문헌고찰은 주로 어떤 보건의료 영역에서 이루어졌나?

IPD 문헌고찰은 1980년대 후반부터 방법론이 꾸준히 발전해 온 심혈관 질환 분야과 암 분야에서 시작되었으며, 현재도 이 분야에서 주로 시행되고 있다. IPD는 그러나 당뇨, 감염, 정신 건강, 치매, 간질, 탈장, 호흡기 질환을 포함한 많은 다른 분야에서도 이루어지고 있다.

18.1.5 IPD 문헌고찰은 언제 이루어지는 것이 적절한가?

일반적으로, IPD 문헌고찰은 이용 가능한 출판물이나 기타 집계 자료에 기초한 양질의 문헌고찰이 이용가능하지 않은 상황에서 고려되어야 한다. 구체적으로는, IPD의 수집이 목표, 자료 양 및 품질, 필요한 분석 등의 측면에서 기존의 집계 자료 접근방식에 비해 어떤 가치를 더 창출할지를 신중하게 고려할 필요가 있다. 이는 첫 단계로서 집계 자료 체계적 문헌고찰(aggregate data systematic review)를 먼저 수행하거나 의뢰할 수 있다는 것을 의미한다. 또는, 핵심 목적이 하위집단과 잠재적 효과변경자(effect modifiers)을 탐색하는 것이라면, IPD 문헌고찰을 직접 진행하는 것이 타당할 수 있다.

신뢰성 있는 분석을 위해서 추가적으로 고려해야 하는 것은 충분한 IPD를 이용할 수 있을 것인가 하는 것이다. 예를 들어, 일부 임상시험 자료는 손실되었을 수 있으며, 유해효과나 삶의 질과 같은 일부 중재결과는 모든 임상시험에서 체계적으로 수집되지 않았을 수도 있고, 또는 임상시험 연구자들이 협력하기를 원하지 않을 수도 있다. 또한, IPD 문헌고찰 결과를 적합한 시간내에 작성하는 것이 불가능할 수 있으며, 어떤 상황에서는 추가로 필요한 자원이 엄청나게 많을 수 있다. 이러한 다양한 요인의 가중치를 잘 고려하여, IPD 문헌고찰이 가장 많은

이익을 가져올 수 있는 시기를 결정해야 한다.

IPD 문헌고찰에 착수하기 전에, 프로젝트가 성공하기 위해 어떤 기술과 자원이 필요할지에 대해 신중하게 생각하고 조언을 구할 필요가 있다.

18.2 IPD 수집

18.2.1 원 연구자로부터 자료 수집

전형적으로 IPD에 기반한 체계적 문헌고찰은 하나 이상의 사전 지정된 문제를 해결하기 위한 국제적인 협력 프로젝트다. 그것은 임상의사와 협력하여 체계적 문헌고찰 저자가 시작하기도 하지만, 임상시험 연구자 콘소시엄이나 자금 제공자의 요청에 따라 이루어지기도 한다.

여러 나라의 다양한 분야의 연구자과 협업을 유지하고 협상하는 데는 상당한 시간과 노력이 들 수 있다. 예를 들어, 적격 연구 책임자를 찾는 것이 어려울 수 있으며, 메타 분석에 참여하는 것을 꺼릴 수도 있다. 흔히 사용되는 첫 번째 접근방식은 주요 연구자에게 이메일이나 편지로 협력을 요청하고, 프로젝트를 설명하며 자료가 필요한 사항과 메타 분석이 어떻게 관리되고 발표되는지를 설명하는 것이다. 프로토콜은 일반적으로 이 단계에서 더 자세한 정보를 제공하기 위해 제공되지만, 대개 자료를 첫 번째 단계부터 찾지는 않는다. 또한 연구 자료의 관리를 책임지고 자료를 보낼 자료 센터 또는 연구 조직과 추가적인 연락을 설정할 필요가 있을 수 있다.

연구자들이 IPD 문헌고찰에 참여하도록 장려함에 있어, 가능한 한 융통성 있게 행동하고, 관계를 구축하는 데 필요한 시간을 갖고, 모든 협력자들이 참여하여 진전을 알 수 있도록 하는 것이 중요하다. 특히 프로젝트가 장기간에 걸쳐 진행될 수 있기 때문에 정기적인 뉴스레터, 이메일 업데이트 또는 웹 사이트가 유용할 수 있다.

18.2.2 원본 연구자 이외의 출처로부터 자료 수집

Yale Open Data, Clinical Study Data Request, DataSphere 또는 Vivli와 같은 일반적인 자료 공유 플랫폼을 이용하면 IPD의 가용성이 높아질 수 있다. 이들은 연방 기관(예: NIH), 자금 제공자(예: MRC), 학술지 편집자, AllTrials campaign 및 코크란 연합에서 임상 연구의 IPD를 더 쉽게 이용할 수 있도록 만들어졌다.

이러한 노력의 초점은 개별 연구 자료를 이용할 수 있도록 하는 것이기 때문에, 포맷과 코딩이 반드시 표준적이거나 다른 연구 자료 집합 전체에 걸쳐 일관적인 것은 아니다. 일부 플랫폼은 IPD에 제한 없이 완전히 접속할 수 있고 다른 플랫폼은 프로젝트 제안서에 대한 승인 여부에 따라 공개될 수 있다. 또한, 리뷰를 실시하는 연구팀에 IPD의 직접 이전을 허용하는 출처도 있지만, 플랫폼 내의 보안 영역 내에서만 IPD를 이용할 수 있는 소스도 있다. 따라서 상당히

복잡한 프로세스가 필요할 수도 있다. 향후 자료 공유 문화가 속도를 내게 되면, IPD의 가용성과 접근성이 증가하게 되고, 결국 IPD 문헌고찰의 생산에 도움이 될 것이다.

18.2.3 '주제 기반' 저장소 구축

범위가 좁은 IPD 문헌고찰 또는 넓은 기반 자료 공유 저장소에 대한 대안으로, 특정 보건의료 영역 또는 주제에 대한 모든 연구로부터 IPD 저장소를 구축할 수 있다. 이전에 이러한 저장소는 기존의 협력 IPD 문헌고찰을 기반으로 구축되었으며 임상 질문을 심층적으로 조사하고 추가 질문을 잠재적으로 처리할 수 있는 고유한 자원으로서의 역할을 담당했다.

1985년 이후 초기 유방암 임상시험자 협력 그룹은 장기적으로는 모든 주요 중재의 영향을 평가하기 위해 초기 유방암에 대한 대다수의 임상시험 자료를 모으고 연장 자료까지 수집했다(http://gas.ndph.ox.ac.uk/ebctcg). 이런 자료를 바탕으로 에스트로겐 수용체 양성 유방암에 걸린 여성들이 내분비 치료 20년 이상 후에도 여전히 상당한 암 재발 위험이 있다는 것을 알게 되었다. 기존 대장암 IPD 문헌고찰을 기반으로 구축된 ACCENT 저장소는 질병 없는 생존을 전체 생존을 대체할 수 있는 중재결과로 확인하였고, 체질량 지수가 생존에 미치는 예측적 영향에 대한 메타분석을 시행하기도 하였다.

이러한 저장소의 주요 장점은 자료 항목을 처음부터 공통 형식으로 코드화 할 수 있어 후속 자료 재사용을 촉진할 수 있으며, 주제에 대한 전문지식을 가진 사람이 IPD를 확인할 수 있다는 것이다. 또한 연구자들과 함께 일함으로써 얻을 수 있는 이익도 유지된다. 물론 IPD의 보유·재이용도 당초의 문헌고찰과 동일한 자료 보안·비밀 보호 대책에 따라야 하며, 필요한 경우 새로운 윤리적 승인, 자료이용 협약을 모색해야 한다. 새로운 분석은 사전 지정된 새로운 프로토콜 및/또는 분석 계획을 따르는 것이 매우 중요하다.

18.2.4 자료 보안 및 기밀성

연구자들은 자연스럽게 그들의 연구 자료가 적절하게 전송되고 저장되고 사용될 수 있도록 하는 안전장치가 있을 것으로 기대한다. 이 때문에, 원연구자와 IPD 문헌고찰팀 간의 자료 공유나 자료 이용 협정이 통상적으로 요구된다. 그러한 계약의 세부사항은 다양하지만, 대부분은 자료를 안전하게 보관하고, 프로젝트 팀의 공인된 구성원만이 접근하며, 다른 곳에 복사하거나 배포하지 않는다고 명시할 것이다. 또한 익명화하거나, 코드화하여 제공된 자료에서 개별 연구대상자의 개인정보를 해제해야 하며, 자료이용 협약을 통해 연구자가 개인을 다시 식별하려는 시도를 금지해야 한다. 필요한 식별 해제 절차는 연구가 시작된 국가의 개인정보 보호 법률에 따라 결정될 수 있다.

역사적으로 IPD 문헌고찰은 윤리적 승인을 받지 않았는데 연구대상자들에 대한 동의를 받은 연구 자료를 사용하기 때문이다. 그러나 진화하는 자료 보호 규정(예: the EU General

Data Protection Regulation), 자료 공유에 대한 태도 변화 때문에 IPD 보유 기관이 공식적인 윤리적 승인을 받아야 하는 것으로 바뀌었다. 따라서 연구팀이 운영하는 윤리위원회 등에서 조사가 필요하며 공식적인 검토가 필요하지 않더라도 자료 제공자에게 이에 대한 서면 확인을 보내는 것이 유용할 수 있다. 그러나 이것이 원래 연구대상자에게 새로운 동의를 받아야 하는 것을 의미하는 것은 아니다. 왜냐하면 자료가 익명화되어 있기 때문이며, 많은 상황에서, 고령의 연구대상자(추적관찰하기 어려움), 말기 환자(많은 연구대상자는 이미 사망하였음)에게 추가로 동의를 받는 것은 거의 불가능할 것이다.

18.2.5 수집할 자료 항목 결정

IPD 문헌고찰을 위해 수집할 자료 항목(또는 변수)을 결정할 때는 계획된 분석을 신중하게 고려하는 것이 현명하다. 이를 통해 분석에 필수적인 정보가 검색되지 않거나 자료가 불필요하게 수집될 가능성을 최소화할 수 있다.

또한 자료의 품질을 최대화하고, 분석을 강화하는 목적에 부합해야 한다. 예를 들어, 연구에 포함된 모든 연구대상자와 중재결과에 대한 자료는 이미 보고된 분석인지 여부에 관계없이 검색되어야 한다. 따라서 자료 수집을 시작하기 전에 연구 프로토콜을 검토하고, 연구자와 접촉하여 실제로 사용할 수 있는 자료가 무엇인지를 확인하는 것이 바람직하다. 많은 경우에, 개별 연구에서 정의된 중재결과 및 연구대상자 특성만 수집하면 된다. 그러나 더 세분화된 중재결과(예: 삶의 질 척도의 하위 척도) 또는 기타 변수를 각 연구에 대해 일관된 방식으로 정의할 수 있도록 추가 변수가 필요할 수 있다. 예를 들어, 공통의 정의에 따라 자간전증(preeclamsia)를 재정의 하려면 수축기/이완기 혈압과 단백뇨에 대한 자료가 필요하다.

IPD는 회복까지 시간, 무발작 시간 또는 사망까지 시간 등과 같은 사건발생 시간 중재결과에 대해 자료를 합성하는 가장 실용적인 방법을 제공한다. 따라서 사건 발생 여부(예: 사망)와 사건 발생 날짜(예: 사망일) 및 사건을 경험하지 못한 사람에 대한 마지막 후속 조치 날짜에 대한 자료를 수집하는 것이 중요하다.

18.2.6 충분한 자료 수집

IPD 문헌고찰에 필요한 모든 자료를 얻는 것이 항상 가능한 것은 아니다. 예를 들어, 임상시험 연구자를 추적할 수 없거나 더 이상 자료에 접근할 수 없기 때문에 모든 관련 임상시험에 대한 IPD를 획득하는 것이 어려울 수 있다. 연구자들이 응하지 않거나 참여를 거부하면 불리한 결과 때문 일 수 있으며, 따라서 그러한 임상시험을 포함하지 않으면 비뚤림이 생길 수 있다. 반면에, 품질이 낮은 임상시험을 제공하기를 꺼려한다면, 오히려 메타 분석을 강건하게 만들 수 있다.

18.3 IPD 관리 및 확인

18.3.1 자료 재코딩 및 재정의

필연적으로, IPD 문헌고찰에 포함된 연구들의 자료를 다른 방법으로 수집하고 정의하게 될 것이다. 그러나 자료 항목을 공통 형식으로 재코딩하는 것은 비교적 간단하며, 예를 들어 병기, 중증도에 대한 정의를 일관된 방식으로 조화시켜 자료 합성을 가능하게 할 수 있다. 따라서 연구자에게 필요한 자료와 안전한 자료 전송 프로세스에 대한 명확한 지침을 제공 할뿐만 아니라 각 변수에 대해 선호하는 자료 형식과 코딩을 제공해야 한다. 물론, 연구자가 이러한 사전 지정된 형식에 따라 자료를 준비하지 않으려 하거나 준비할 수 없는 경우, 문헌고찰 팀은 어떤 형식이든 가장 편리한 형식으로 자료를 받고 필요에 따라 재코딩해야 한다. 제공된 자료 복사본은 자료의 변환 또는 수정을 수행하기 전에 보관해야 하며, 변경한 내용이 적절히 기록되어야 한다.

18.3.2 받은 자료의 완전성 및 무결성 확인

자료를 확인하고 '정리'하는 것의 목적은 포함된 자료가 정확하고 유효하며 내부적으로 일관성이 있는지 확인하는 것이다. 문헌고찰 팀이 자료를 독립적으로 조사하여, 자료의 신뢰도를 높여야 한다. 자료 파일을 처음 받았을 때, 해당 파일을 읽고 중앙 저장/분석 시스템에 업로드 할 수 있는지 확인하는 것이 중요하다. 예를 들어 전자 자료인 경우, 파일을 열 수 있고 자료가 올바른지 확인해야 한다. 또한 모집 또는 무작위배정 된 모든 연구대상자가 포함되며, 환자 식별자 목록에서 명백한 누락이나 중복이 없음을 확인하는 것이 유용하다.

또한 제공된 자료는 관련 연구 논문 또는 결과 저장소에 대해 확인하여 기저 상태 특성의 분포, 연구대상자 수 및 중재결과와 불일치가 있는지 확인해야 한다. 하지만 등록이 계속되거나 출판 이후의 후속 조치 때문에 차이가 발생할 수 있음을 명심해야 한다.

18.3.3 포함된 연구의 비뚤림위험 확인

다른 유형의 체계적 문헌고찰과 마찬가지로, IPD 문헌고찰에 포함된 연구의 비뚤림위험을 평가해야 한다. IPD 접근법의 경우 연구자가 제공한 프로토콜, 코드북 및 양식에서 얻은 추가 정보로 비뚤림위험 평가의 명확성을 높일 수 있다.

18.3.3.1 무작위배정 및 배정 은폐 확인

무작위배정 임상시험의 경우, IPD를 확인하여 무작위배정되지 않은 연구나 연구대상자가 포함되는 것을 방지하기 위해 무작위배정 및 배정 은폐 방법이 적절하게 시행되었는지 확인하는 것이 중요하다.

18.3.3.2 탈락 확인

가능한 한 많은 무작위 연구대상자에 대한 자료가 각 중재결과에 대해 포함되었는지, 그리고 배정된 중재에 계속 있었는지를 확인해야 한다. 이는 연구대상자의 탈락 또는 연구 분석에서 제외되는 것과 관련된 비뚤림을 최소화하는 데 도움이 되며. 프로토콜대로 분석의 잠재적 비뚤림을 피할 수 있게 해준다.

18.3.3.3 포함된 중재결과 확인

IPD 문헌고찰에서는 보고하였는지 여부와 관계없이 핵심질문과 관련된 모든 중재결과를 수집해야 한다. 이는 중재결과의 차등 보고와 관련될 수 있는 비뚤림을 극복하는 데 도움이 된다. 정확하게는 일부 측정된 중재결과가 보고되지 않을 수 있기 때문에 연구 프로토콜, 시험 레지스트리 항목 및 연구자와 함께 어떤 결과가 이용 가능한지를 확인할 가치가 있다.

18.4 IPD 분석

18.4.1 장점분석

각 연구에 대해 IPD를 액세스 할 수 있으면 분석 가정을 확인하고 자료를 철저히 탐색하며 여러 임상시험들을 일관되게 분석 할 수 있다. 또한, 위험과 중재결과에 대한 효과 측정치가 IPD 분석에서 직접 도출되므로, 논문에 제시된 정보 및 분석에 의존하거나 다른 방식으로 분석된 연구의 요약 통계를 결합 할 필요가 없다. IPD의 재분석은 원래 분석의 문제점이나 제한을 피하게 해주어 원래 논문에서 수행되지 않았더라도 보다 적절한 효과 측정치를 사용하고 결측 자료를 설명하기 위해 정교한 분석을 수행하게 해준다.

18.4.2 중재의 전반적인 효과 평가

IPD 문헌고찰 연구대상자를 적절히 계층화하거나 설명하는 것이 중요한데 그 이유는 연구대상자는 다른 연구 프로토콜에 따라 모집되었기 때문이다. 단일 '메가'임상시험 IPD를 결합하면 중재의 효과에 대한 비뚤림과 과대추정이 가능하다. 현재까지 대부분의 IPD 메타 분석은 2 단계 분석 방법을 사용하여 각 개별 연구를 첫 단계에서 독립적으로 분석하고, 두 번째 단계에서는 기존의 체계적 문헌고찰과 거의 동일한 방식으로 합성추정치를 제공한다.

18.4.3 시험 특성에 따라 효과가 달라지는지 평가

연구 특성에 따라 중재 효과가 달라지는지 여부를 탐구하는 것은 메타 분석의 중요한 측면이며, 집계 자료에 사용되는 것과 동일한 분석 접근법을 사용하여 IPD로 쉽게 조사할 수 있

다. 따라서 하위군 분석을 사용할 수 있으며, 여기서 연구는 약물 유형과 같은 특정 특성과 이들 그룹 간의 간접 비교 효과에 따라 분류된다. 또는 메타 회귀 분석은 약물 투여량과 같은 연구 치료 특성과 관련하여 중재의 전체적인 효과가 달라지는지 여부를 탐구하는 데 사용될 수 있다.

18.4.4 연구대상자의 특성에 따라 효과가 달라지는지 평가

IPD 문헌고찰은 연구대상자의 특징에 따라 중재 효과가 달라지는지 여부를 가장 신뢰할 수 있고 종종 유일한 방법이다. 예를 들어, 특정 중재가 남성보다 여성에게 더 효과적인지 여부를 볼 수 있다. 이런 분석은 두 단계로 이루어질 수 있다. 첫 번째 단계에서, 성별과 개별 연구대상자 수준의 중재 효과 사이의 상호작용이 각 연구에서 추정되며, 두 번째 단계에서 이러한 상호작용이 표준 메타 분석 기법, 이른바 '임상시험내(within trial)' 상호작용을 사용하여 여러 연구에서 통합된다. 널리 사용되는 '하위군 분석' 접근방식은 우선 각 연구는 하위군으로 구분되며, 남성에서의 효과 메타 분석을 여성에서의 메타 분석과 비교된다. 불행하게도 이 접근방식은 시험내 상호작용을 시험간 상호작용으로 확장하여 비뚤림에 취약하므로 피하는 것이 좋다. 또는 1단계 접근법을 사용할 수 있지만 비뚤림을 피하기 위해 연구 내 상호작용과 연구 간 상호작용을 구별하는 데 주의를 기울여야 한다.

18.5 IPD 문헌고찰의 품질 평가

분명히 상당한 장점을 제공하며, 다양한 보건의료 영역에서 사용이 증가하고 있지만 모든 IPD 문헌고찰이 동일한 표준에 수행되거나 보고되는 것은 아니다. 더욱이 IPD의 수집, 확인, 분석 과정은 집계 자료보다 더 복잡하며, 일반적으로 보고해야 할 분석이 더 많기 때문에 IPD 문헌고찰의 품질을 판단하기 어려울 수 있다. 이는 다시 정책, 임상진료 및 새로운 시험에 영향을 미칠 수 있는 보급을 방해할 수 있다.

18.6 참고문헌과 읽을거리

- Debray TP, Riley RD, Rovers MM, Reitsma JB, Moons KG, Cochrane IPD Meta-analysis Methods group. Individual participant data(IPD) meta-analyses of diagnostic and prognostic modeling studies: guidance on their use. PLoS Medicine 2015a; 12: e1001886.
- Stewart GB, Altman DG, Askie LM, Duley L, Simmonds MC, Stewart LA. Statistical analysis of individual participant data meta-analyses: a comparison of methods and recommendations for practice. PloS One 2012; 7: e46042.

- Tierney JF, Pignon J-P, Gueffyier F, Clarke M, Askie L, Vale CL, Burdett S. How individual participant data meta-analyses can influence trial design and conduct Journal of Clinical Epidemiology 2015b; 68: 1325-1335.
- Tierney JF, Stewart LA, Clarke M. Chapter 26: Individual participant data. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.0 (updated July 2019). Cochrane, 2019. Available from www.training.cochrane.org/handbook.

제19장 기존의 체계적 문헌고찰의 활용 (Using Existing Systematic Reviews)

핵심 포인트

- 기존의 체계적 문헌고찰을 활용하는 것이 효율적일 수 있으나, 아직 방법론적으로 정립되지 않았으므로 장점과 단점을 잘 검토하여 활용한다.
- 연구 진행 순서는 1) 사전에 정한 PICOTS-SD에 맞는 체계적 문헌고찰이 있는지 검색, 2) 적절성 평가, 3) AMSTAR를 이용한 체계적 문헌고찰 질평가 수행, 4) 활용방법 결정, 5) 자료 추출 및 이용의 단계를 거치게 된다.
- 검색 시점, 자료이용 합성 방법, 체계적 문헌고찰의 최선 활용 방법, 추가적 근거의 결합 방법 등에 이슈가 있으며, 아직 정립중인 방법론이기 때문에 실제로 이용할 때는 많은 검토가 필요하다.

연구에 따라서는 새로운 체계적 문헌고찰 연구를 수행하기보다는, 기존에 출판된 체계적 문헌고찰을 활용하여 의료기술평가(Health technology assessment, HTA)를 시행하거나 임상진료지침(Clinical practice guideline, CPG)을 개발하는 것이 효율적일 수 있다. 이러한 연구방법은 아직 방법론적으로 정립되어 있지는 않으며, 추후 확립될 가능성이 있다. 현재로서는 장점과 단점을 검토하여 사용여부를 결정하여야 한다. 장점과 단점은 아래와 같다.

표 19-1. 기존의 체계적 문헌고찰 활용

장점	• 시간, 노력, 자원을 아낄 수 있음 • 누락된 연구가 있는지 중복 확인할 수 있음 • 연구 중에 생길 수 있는 방법론적 이슈를 예측, 계획할 수 있음 • 불필요한 중복 연구를 피할 수 있음 • 현재 접근할 수 없는 연구나 구할 수 없는 정보를 검토할 수 있음
단점	• 기존 연구가 오래되거나 질이 낮거나, 핵심 질문에 맞는 연구가 없는 경우 시간과 자원을 오히려 낭비할 가능성이 있음 • 기존 연구에 오류가 있는 경우 오류전파 가능성이 있음 • 기존의 연구를 이용하여 결론을 도출하므로 연구를 수행하는 기관의 신뢰성이나 성실성이 떨어지는 것으로 평가 받을 수 있음 • 연구방법에 아직 논란이 있음(특히 기존의 체계적 문헌고찰끼리 비교 하는 방법)

연구수행 방법과 전체 흐름도는 그림 19-1과 같다.

1단계 : 사전에 정한 PICOTS-SD에 맞는 체계적 문헌고찰이 있는지 검색한다.

2단계 : 검색된 체계적 문헌고찰이 적절한지 평가한다.

3단계 : 적절하다고 평가하였다면, 검색된 체계적 문헌고찰에 대한 질평가를 수행한 후 기존의 체계적 문헌고찰을 활용하여 평가할지 결정한다.

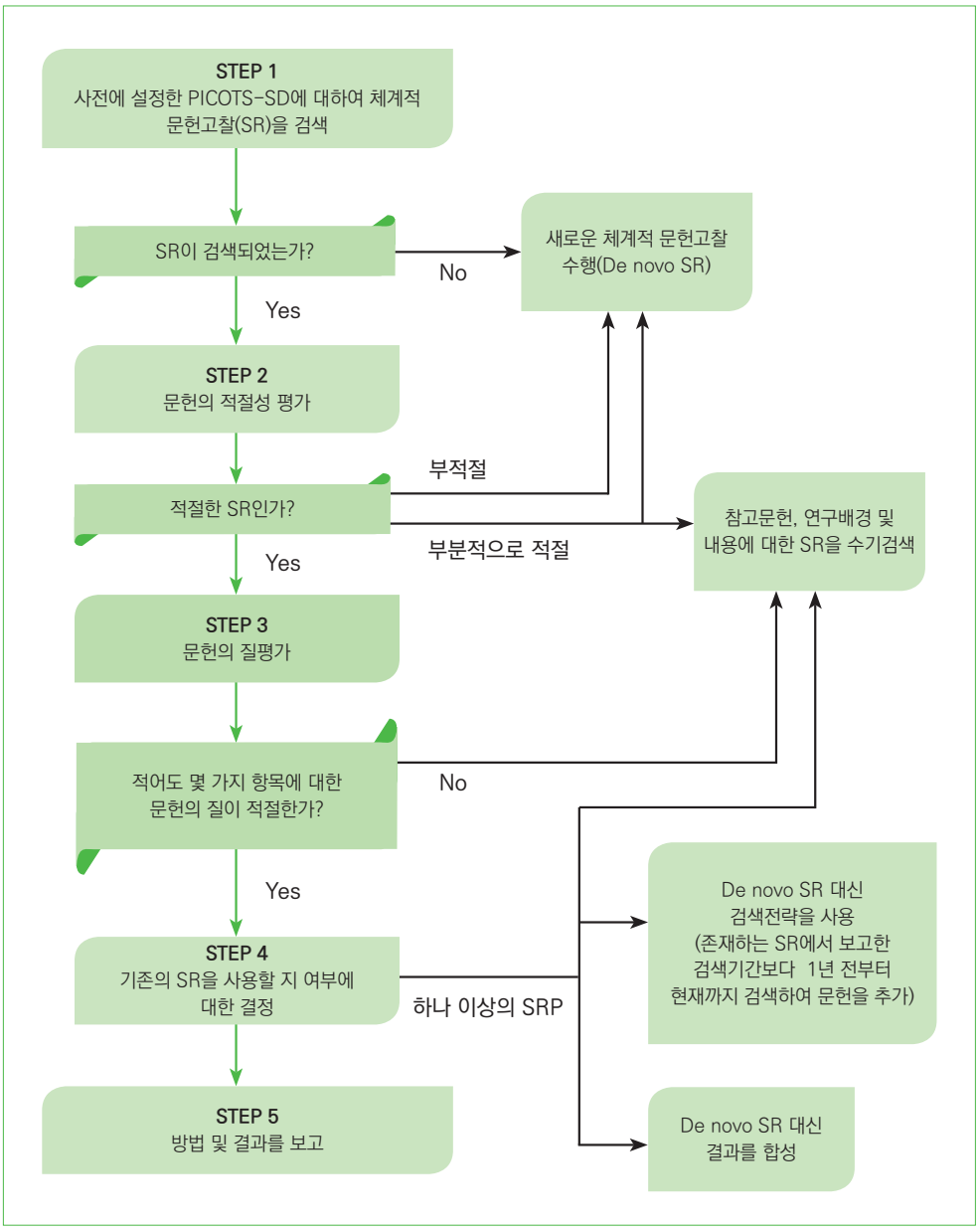


그림 19-1. 기존의 체계적 문헌고찰을 활용한 의료기술평가

기존의 체계적 문헌고찰을 활용한 의료기술평가가 잘 시행될 수 있는 전제조건은 다음과 같다.

- 해당 연구의 해당 PICO를 모두 포함하는 적절한 범위의 체계적 문헌고찰이 존재하는 경우
- AMSTAR 평가 방법론에서 일정 수준 이상의 질을 가진 체계적 문헌고찰이 존재하는 경우

19.1 문헌검색

일반적인 검색방법은 앞 장에서 설명된 문헌검색 부분을 참고하면 되며 여기서는 기존의 체계적 문헌고찰을 활용할 때 적합한 검색 전략만을 제시하고자 한다.

기존의 체계적 문헌고찰을 검색할 때에는 PICOTS-SD에 맞는 검색을 시행해야 한다. 체계적 문헌고찰을 활용할 경우 관련 체계적 문헌고찰을 모두 찾는 것보다는 질 높은 체계적 문헌고찰을 찾는 것이 일차적인 목표가 되기 때문에 많은 문헌을 선별할 수 있는 데이터베이스를 대상으로 검색하는 것이 필요하다. 그러한 데이터베이스에는 Evidence-based Practice Center program, MEDLINE's Top 120 Index Medicus Journals, Health Technology Assessments, The Cochrane Library, DARE 등이 있다.

일반 데이터베이스를 이용하는 경우 체계적 문헌고찰에 특화된 검색전략을 이용할 수 있다. 검증된 검색 필터 목록은 https://sites.google.com/a/york.ac.uk/issg-search-filters-resource/filters-to-identify-systematic-reviews/filters-to-identify-systematic-review-methodology 중 'systematic reviews'에서 확인할 수 있다.

19.2 체계적 문헌고찰의 질평가

체계적 문헌고찰의 질평가는 AMSTAR(assessment of multiple systematic reviews)를 사용한다. 체계적 문헌고찰의 질평가는 잘 수행된 체계적 문헌고찰을 선택하는 중요한 과정이다. 최근에 개발된 AMSTAR2나 ROBIS를 이용해도 된다.

사용하는 질평가도구	적용 연구설계
AMSTAR	체계적 문헌고찰(Systematic reviews)

체계적 문헌고찰의 질평가는 최소한 2명 이상이 수행하여야 하며 의견이 다를 때 이를 해소할 수 있는 방법을 명시한다. 한 두 연구를 대상으로 시범 평가를 수행하여 평가자의 의견이 일치하는 지 검증한다.

AMSATR는 총 11개 항목을 평가하도록 이루어진 체크리스트 방식의 도구이며, 타당도 검

증(validation)이 된 도구라는 장점을 지닌다(Shea, 2009). 평가자는 도구를 이용할 때 각 부분 항목에 대하여 '예', '아니요', '대답할 수 없음', '적용할 수 없음' 네 가지로 대답한다. '예'는 검토하는 체계적 문헌고찰이 해당항목에 관하여 적절히 다루었다는 것을 의미하며, '아니요'는 해당항목에 관하여 잘 수행되지 않았을 가능성을 의미한다. '대답할 수 없음'은 해당항목을 시행했는지 알 수 없는 경우를 의미하며, '적용할 수 없음'은 해당항목을 적용할 수 없어 수행하지 못한 경우라 할 수 있다.

하지만, 아직까지는 AMSTAR 질평가 결과를 어떻게 활용할 것인지에 대해 합의가 되지 않은 부분이 있다. AMSTAR에서 주요 영역을 정하고 주요 영역의 문항에 가중치를 부여할지에 대한 여부와 체계적 문헌고찰의 질을 범주화 하는 방법이다. AMSTAR 결과를 이용하는 방법은 다양하고, 합의된 방법이 없는 만큼 어떠한 방법을 선택할 것인지 연구진 내부의 충분한 논의가 필요할 수 있다. 현재로서는 AMSTAR를 점수화하여 0-3점이면 낮은 질, 4-7점이면 중등도의 질, 8-11점은 높은 질로 분류하는 방법이 추천된다.

AMSTAR로 체계적 문헌고찰의 질을 평가하여 일정 정도의 질을 가진 것으로 판단될 때, 해당 체계적 문헌고찰을 활용할 수 있다. 일정 정도는 현재 합의된 내용이 없으므로 연구진에서 합의하여 결정한다. 일정 정도 이상의 질을 가진 체계적 문헌고찰이 없는 경우에는 새로운 체계적 문헌고찰을 시행할 수 있다.

미국 공중보건서비스 특별위원회에서는 기존 체계적문헌고찰 활용을 위한 고려 사항을 정하였으며, 이를 체크리스트 형태로 제시하고 있으며 그 내용은 아래와 같다.

- 리뷰가 수용할 만한 체계적 검색을 수행하였나?
- 리뷰는 중재들의 효과, 이득에 대한 정보를 포함하는가?
- 질평가가 된 연구들이 확인되고 자세히 요약되었는가?
- 질평가된 연구 근거체는 많은 수의 연구들을 포함하는가?
- 중재를 다루는 유일한 연구인가?최근에 수행된 연구인가?
- 이 리뷰는 불충분한 근거가 아니고, 효과가 있다는 결론을 제공하는가?
- 검색 종료일이 최근인가?
- 연구는 질평가에 따라 포함 배제 되었는가?
- 리뷰는 권고를 위한 근거들의 효과측정에 대한 적절한 정보, 요약을 제공하는가?
- 양질의 연구들의 결론에서 일관성 있는 결정을 하기에 적절한 정보를 제공하는가?
- 제공된 정보가 효과의 크기가 충분한지 여부를 결정하기에 적절한가?
- 정보는 적절한 세팅, 대상, 중재 제공자, 중재 특성에서 일반화하기에 적절한가?
- 질 평가된 근거를 종합하여 효과를 결론 내릴 수 있었는가?

미국 공중보건서비스 특별위원회에서는 질평가와 함께 최신성, 적절성 등을 종합적으로 평

가하여 기존의 체계적 문헌고찰의 활용에 대한 기준으로 삼고 있다(그림 19-2)

19.3 자료추출

체계적 문헌고찰의 어떠한 내용을 어떻게 이용할지는 매우 다양한 방법이 있다. 아래 표에 제시된 목록 중 필요한 항목을 선정하여 해당 내용을 체계적 문헌고찰에서 추출한다.

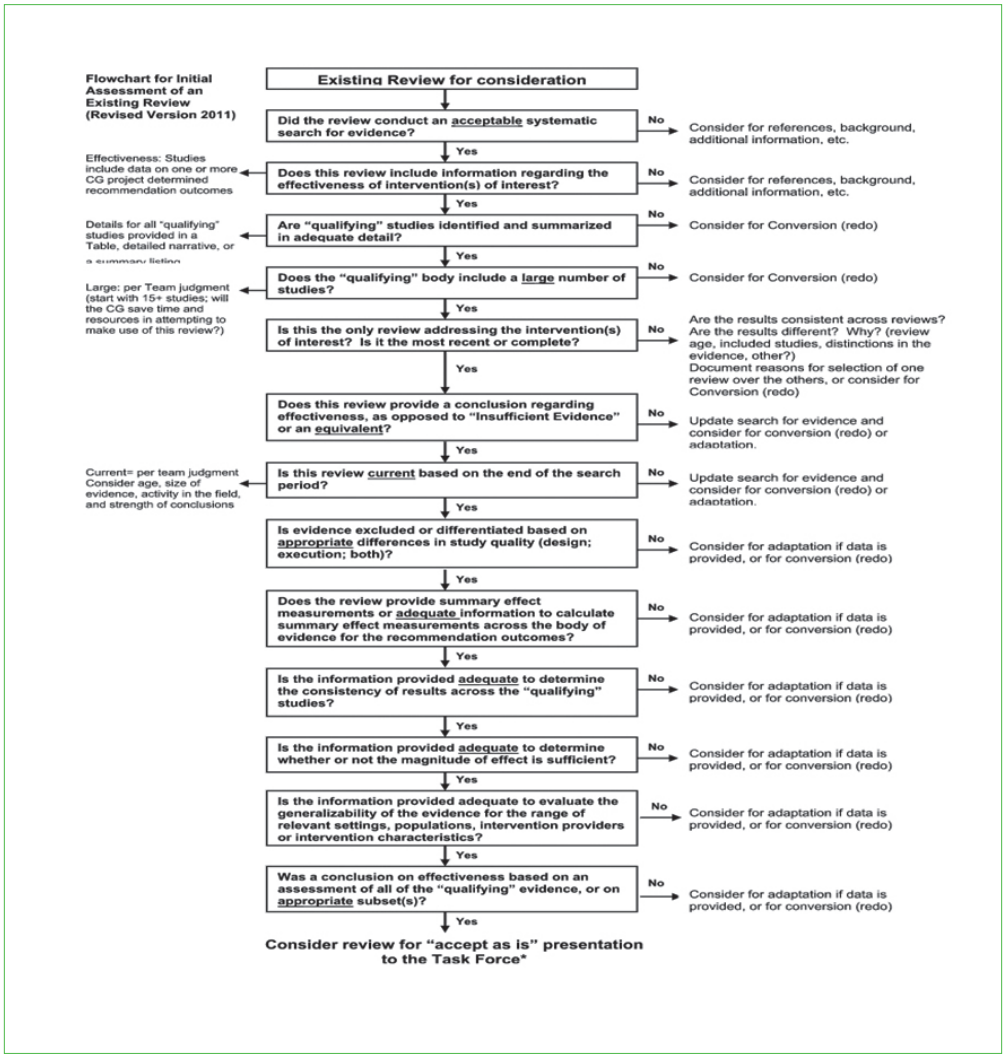


그림 19-2. 기존 체계적문헌고찰 활용 방법 결정을 위한 알고리듬(CPSTF).

표 19-2. 선택된 체계적 문헌고찰에서 이용 가능한 자료 목록.(AHRQ, 2009) Adapted table AHRQ(Original publisher)

- 일차 연구 목록
- 일차 연구의 추출값
- 질평가 도구 및 결과*
- 체계적 문헌고찰의 결론
- 검색 전략
- 포함/배제 기준

* 단, 다른 질 평가 도구를 이용하여 평가할 예정이라면 일차문헌을 모두 찾아 재평가해야 한다.

19.4 추가검색

체계적 문헌고찰의 주제가 임상연구가 많을 가능성이 있는 분야인 경우에는 추가 검색이 필요할 수 있다. 선택한 체계적 문헌고찰의 최종 검색일을 확인한 후, 최종검색일의 1년 전부터 현재까지 검색한다. 일차 문헌의 검색 방법은 'PART I. 3. 검색방법'을 참고하면 된다. 검색 전략의 경우 선택한 체계적 문헌고찰에 포함되어 있다면, 그대로 이용할 수도 있으며, 새롭게 개발할 수도 있다. 추가 검색된 문헌이 있다면, 질평가와 자료추출, 합성 등을 새로 시행할 필요가 있다.

19.5 기존 체계적 문헌고찰 이용 방법

선택된 체계적 문헌고찰의 결과를 어떻게 이용할지에 대해서도 아래와 같은 다양한 방법들이 있다.

- 기존 체계적 문헌고찰의 결과를 요약하여 결론 자체로 받아들임
- 최신 근거를 추가하여 메타분석하고 이를 바탕으로 기존의 체계적 문헌고찰 재구성함
- 문헌 검색 전략을 이용하여 추가 검색하고 새로운 체계적 문헌고찰 수행함
- 여러 체계적 문헌고찰이 있을 때 결과가 서로 다른 부분의 이유를 다룸
- 문헌 검색 전략만을 받아들임

19.6 방법과 결과의 보고

기존의 체계적 문헌고찰을 활용한 체계적 문헌고찰을 수행하면서 사용한 방법과 결과를 아래 그림 19-2와 같은 형태로 보고할 수 있다.

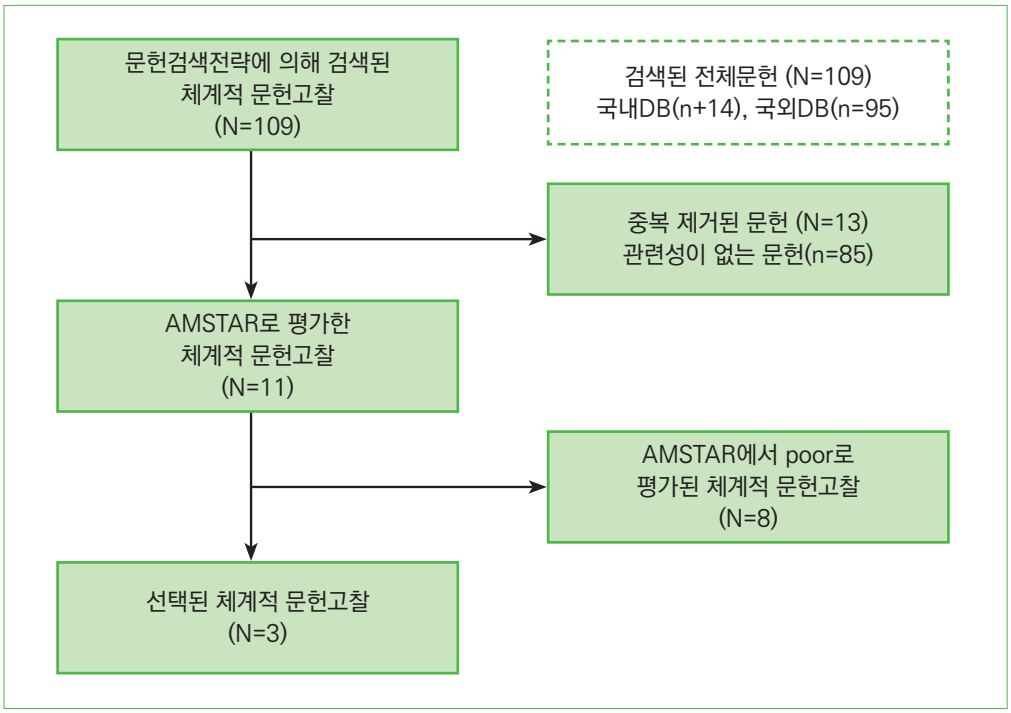


그림 19-3. 기존의 체계적 문헌고찰 이용 연구 흐름도

19.7 관련 이슈

기존의 체계적 문헌고찰을 활용하는데 있어 아직까지 합의된 연구방법이 없기 때문에 많은 이슈가 있을 수 있다.

아직까지 합의된 연구방법이 없기 때문에 많은 이슈가 있을 수 있다.

- 검색 시점은?

최근의 체계적 문헌고찰을 이용할 가능성이 높기 때문에 언제 이후의 체계적 문헌고찰을 검색할 것인지 논란이 있을 수 있다. 일반적으로 2000년 이후에 체계적 문헌고찰의 출판이 활발해졌기 때문에, 그 이후의 연구를 검색할 수 있겠다. 중재 등의 도입 시기를 고려하여 년도를 정할 수 있다.

- 자료이용 합성 방법

3절과 5절에서 논의한 것처럼 자료를 이용하거나 합성하는 방법은 매우 다양할 수 있다. 결론을 그대로 이용하느냐, 추출된 자료를 이용하여 새로운 합성을 시도하느냐 등은 주제 및 연구현황에 따라 달라질 수 있기 때문에 어떠한 방법을 선택할 것인지 많은 검토가 필요하다.

• 체계적 문헌고찰에 대한 최선의 활용 방법

아직까지 어떻게 활용하여 연구결과를 도출하는 것이 최선인지 연구된 바가 없기 때문에 다양한 방법으로 활용되고 있다. 이에 대한 추가적인 연구가 필요하다.

• 기존 체계적 문헌고찰에 추가적인 근거를 결합하는 방법

추가검색을 통해 새롭게 포함된 문헌이 있다면, 이를 어떻게 기존의 연구결과에 합성할지에 대해 논의가 필요하다.

19.8 참고문헌과 읽을거리

- Shea B, Grimshaw J, Wells G, Boers M, Andersson N, Hamel C, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC medical research methodology. 2007;7(1):10-17.
- Shea B, Hamel C, Wells G, Bouter L, Kristjansson E, Grimshaw J, et al. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. Journal of clinical epidemiology. 2009;62(10):1013-20.
- White CM, Ip S, McPheeters M, et al. Using existing systematic reviews to replace de novo processes in conducting Comparative Effectiveness Reviews. In: Agency for Healthcare Research and Quality. Methods Guide for Comparative Effectiveness Reviews [posted September 2009]. Rockville, MD. available at: http://effectivehealthcare.ahrq.gov/healthInfo.cfm?infotype=rr&ProcessID=60.
- Whitlock EP, Lin JS, Chou R, Shekelle P, Robinson KA. Using existing systematic reviews in complex systematic reviews. Ann Intern Med 2008;148:776-82.

제20장 신속 체계적 문헌고찰

핵심 포인트

- 신속 체계적 문헌고찰은 표준적 체계적 문헌고찰 방법을 신속하게 진행하거나 일부 과정을 생략하여 연구진행을 4주-16주 이내에 완료하는 것으로, 대부분 정책결정자의 필요성에 의해 수행된다.
- 신속 체계적 문헌고찰 요청자가 초기부터 지속적으로 관여하는 것이 신속 체계적 문헌고찰이 요청자의 요구에 적합한지 확인하는데 필수적이다. 통상 문헌고찰의 프로토콜은 시작시기에 작성되지만 신속 체계적 문헌고찰을 위한 프로토콜은 이해당사자와 협의하여 수정할 수 있다. 프로토콜에 대한 변경 사항도 최종 보고서에 반영되어야 한다.
- 신속 체계적 문헌고찰의 과정은 다음과 같다: 1) 신속 체계적 문헌고찰 필요성 평가, 2) 주제 결정/ 핵심질문 개발, 3) 프로토콜 개발, 4) 검색, 문헌 선정, 5) 자료 추출, 6) 자료의 합성, 7) 보고서 작성, 8) 추후관리이다. 연구수행과정에 정책결정자(의뢰자)가 주로 참여하는 단계는 필요성 평가, 주제 결정, 추후관리이다.
- 날짜와 언어로 검색을 제한하고, 검색된 전자 데이터베이스의 수를 제한하며, 한명의 연구자가 문헌선택, 비뚤림위험평가, 자료추출 하는 등 검색에서 합성에 이르기까지 모든 문헌고찰 과정 단계에서 방법을 간소화할 수 있다.
- 연구자는 문헌고찰이 의도한 목적에 적합하도록 이해 관계자의 의견을 바탕으로 투명한 방법론적 선택을 해야 하며, 모든 내용은 투명성의 원칙으로 관리해야 한다.

20.1 개요

보건정책 및 시스템에 정보를 제공하기 위해 신속 체계적 문헌고찰의 필요성이 대두되고 있다. 정책 입안자와 건강 시스템 관리자는 보건 시스템 중재 및 정책의 효과를 포함하여 이러한 중재의 작동 방식 및 세팅, 비용 효과성 등 다양한 문제를 다루는 신속 체계적 문헌고찰을 요구한다. 체계적 문헌고찰은 정책결정에 정보를 제공하기 위해 점점 사용이 증가되고 있지만, 종종 의사결정 시간이 길어지거나 기간 내에 보고서가 생산되지 못하는 경우가 있다. 이러한 문제로 인하여, 신속 체계적 문헌고찰은 응급상황뿐만 아니라 일상적인 보건시스템에 대한 정보에 입각한 결정을 내릴 수 있는 실행 가능한 관련 근거를 신속하게 제공하는 유용한 접근법이다. 따라서 신속 체계적 문헌고찰은 지식 합성의 핵심원칙을 존중하는 투명하고, 과학적이며, 재현 가능한 방법을 통해 생산되어야 한다. 신속 체계적 문헌고찰을 수행하는 다양한 방법

이 존재하며, 주요 과제는 엄격성과 투명성을 유지하면서 검토 방법을 가속화하는 데 있다.

협의의 신속 체계적 문헌고찰은 체계적 문헌고찰 방법을 빨리 진행하거나 일부 생략된 방법으로 4-16주에 걸쳐 진행하는 것으로, 일반적으로 짧은 시간 내에 정보를 생성하기 위해 체계적 문헌고찰의 구성 요소를 단순화하거나 생략하는 일종의 지식 합성 유형으로 정의된다.

신속 체계적 문헌고찰의 목적은 표준화되고 투명한 방법론을 사용하되 효율성을 높이는 것이다. 신속 체계적 문헌고찰의 경우 대부분 정책결정자라는 주요한 독자를 가진다. 따라서 과정은 정책결정자의 필요성에 맞추어야 하며 이러한 목적에 충실해야 한다. 신속 체계적 문헌고찰을 할 때는 기존의 체계적 문헌고찰이 있는 경우 이를 활용할 수도 있고, 기존 체계적 문헌고찰을 갱신할 수 있으며, 새로운 체계적 문헌고찰을 수행할 수도 있다. 체계적 문헌고찰 단계에서 생략하여 신속 체계적 문헌고찰을 할 수 있는 방법은 [표 20-1]과 같다.

표 20-1. 신속 체계적문헌고찰에서 가능한 방법과 타당도에 미치는 영향

체계적 문헌고찰 단계	가능한 방법	타당도에 미치는 영향
프로토콜 작성	프로토콜 생략	알려지지 않음
질문 생성	질문의 수, 하부 질문의 수를 제한, 질문의 범위를 제한	없음
문헌검색	DB 수 제한, 참고문헌 검색 제한, 추가 논문 확인 위한 자문 제한	DB 수를 줄이면 효율은 높이지만 타당성은 줄어들 수 있음.
논문 선택	한 명 제목/초록만, 한 명 전문 선별, 한 명 제외자료 검토	알려지지 않음. 한 명이 9% 정도 놓친다는 보고 있음
자료 추출	한 명 자료 추출/한 명은 점검/ 자료추출은 핵심적인 내용만	오류 증가할 수 있지만 정확한 영향은 알려지지 않음
포함기준	회색문헌 제외	출판비뚤림 가능하지만 근거는 mixed
	언어제한(영어만 포함)	질문에 따라 다름
	최근 10년간처럼 기간 제한	예상되지 않음
	체계적 문헌고찰, RCT 등 일부 연구 설계만 포함	예상되지 않음
질평가	질평가 제한하거나 생략	권장되지 않음
자료 합성	사전 계획 생략	알려지지 않음
	한 명이 시행	알려지지 않음
	메타분석 하지 않음	

체계적 문헌고찰 단계	가능한 방법	타당도에 미치는 영향
출판비뚤림 평가	생략	알려지지 않음
이해상충 평가	생략	알려지지 않음
보고	간소화 혹은 생략	알려지지 않음
외부 검토	간략화 혹은 생략	알려지지 않음

20.2 신속 체계적 문헌고찰의 과정

- 신속 체계적 문헌고찰의 과정은 다음과 같다: 1) 신속 체계적 문헌고찰 필요성 평가, 2) 주제 결정/ 핵심질문 개발, 3) 프로토콜 개발, 4) 검색, 문헌 선정, 5) 자료 추출, 6) 자료의 합성, 7) 보고서 작성, 8) 추후관리이다. 연구수행과정에 정책결정자(의뢰자)가 주로 참여하는 단계는 필요성 평가, 주제 결정, 추후관리이다.

20.2.1 필요성 평가

의뢰자 혹은 문헌고찰 사용자(의뢰자)와 회의를 통해 ① 질문의 범위, ② 문헌고찰의 목적, ③ 시간표, ④ 의뢰자의 책임성과 문헌고찰의 개입 정도 등에 대해 논의한다.

이 과정을 통해 신속 체계적 문헌고찰의 범위는 무엇이고, 신속고찰 방법론이 원래 의뢰자의 문헌고찰 목적에 부합하는지, 그리고 신속문헌고찰을 통해 산출되는 최종 결과물이 의뢰자에게 의미가 있는지를 확인하게 해야 한다.

기존의 체계적 문헌고찰 방법론이 아닌 신속 체계적 문헌고찰 접근법이 적합한지에 대한 내부 평가서를 만드는 것이 좋다. 즉 요청한 내용, 의뢰자 분류(중앙 정부기관, 병원, 지방 정부기구 등), 신속문헌고찰을 의뢰한 이유(시간적 긴박성, 제한된 시간이나 자원 때문) 등의 내용을 포함한다.

이때 결정해야 하는 것은 신속문헌고찰이 필요한가이다. 신속 체계적 문헌고찰의 경우 일정 정도의 과학적 엄격성을 희생하면서 효율성을 추구하는 것이기 때문이다. 따라서 신속 체계적 문헌고찰은 가능하면 명백한 시간적 제한이 있는 경우만 선택하도록 하여야 한다(그림 20-1).

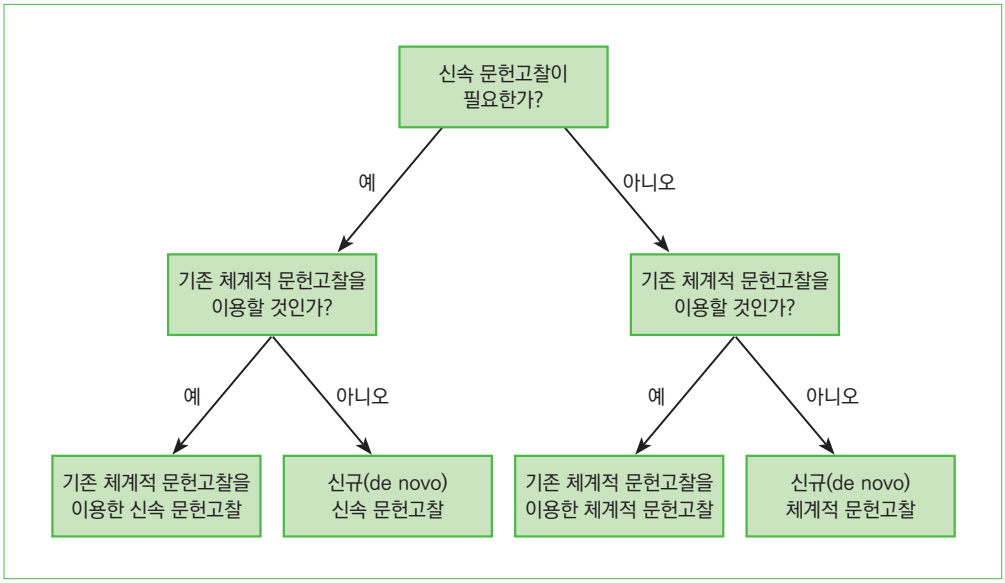


그림 20-1. 체계적 문헌고찰 필요성과 방법 결정 알고리즘

20.2.2 주제 결정/ 핵심질문 개발

주제 결정 단계는 가장 중요한 단계라고 할 수 있으며, 이후 작업의 틀을 마련하는 단계이다. 의뢰자에게 억지로 맞추기 보다는 적절한 연구 질문을 만드는 것이 좋다. 의뢰자 혹은 예상 정보 사용자가 알고 싶은 내용은 광범위할 가능성이 많으나, 이를 토대로 정밀하고 대답 가능한 질문을 만드는 것이 과제이다. 1-2일 정도 해당 주제에 대한 깊이 있는 탐구를 하는 것이 필요하다. 문헌고찰의 범위는 한 번에 결정되지 않을 수 있고, 일단 결정한 이후 추후에 수정할 수도 있다(질문, 근거의 종류, 일차 중재결과 등). 범위가 결정 되었으면 이를 바탕으로 핵심질문의 리스트를 만든다. 가능한 핵심 질문의 예시는 [표 20-2]와 같다.

표 20-2. 신속 체계적 문헌고찰 핵심 질문 예시

- 특정 인구집단에서 특정 중재법의 이득과 유해는?
- 새로운 서비스를 제공하고 구성하는 것의 이득과 유해는?
- 질향상 프로그램의 이득과 유해는?
- 진단검사 X의 정확도는?
- X 중재법을 받는 지역주민의 환자 경험은 어떠한가?
- X 질병의 유병률은?
- 유전자 X와 질병 X 사이의 연관성은 얼마나 강한가?

핵심질문은 PICOT-SD(population, intervention, comparison, outcome, timing, study design) 포맷으로 정리하는 것이 좋다 .

20.2.3 제 3 단계 프로토콜 개발

프로토콜은 연구 계획을 명시하는 문서이며, 중요한 품질 관리 도구이다. 이는 프로세스의 투명성 확보에도 중요한 역할을 한다. 하지만 필요에 따라 신속 체계적 문헌고찰 과정에서 프로토콜을 생략하는 경우도 있고, 이러한 방법론이 타당도에 어떠한 영향을 미치는지는 알려져 있지 않다. 하지만 과정의 투명성과 체계화 측면에서 프로토콜을 생략하는 것은 추천되지 않는다.

프로토콜에는 [표 20-3]와 같은 내용이 포함될 수 있다.

표 20-3. 신속 체계적 문헌고찰 프로토콜의 내용

- 배경
- 연구 질문
- 연구 방법
- 시간표
- 의사결정자의 역할

중요한 결정 중의 하나는 기존 체계적 문헌고찰을 이용할 것인가이다. 기존 체계적 문헌고찰이 있는 경우 이를 이용하여 문헌고찰을 진행할 수 있다. 하지만 다른 연구자들이 시행한 체계적 문헌고찰의 타당도에 대한 신뢰가 있어야 가능하다(그림 20-1).

20.2.4 문헌 검색

체계적인 검색은 문헌고찰의 타당성에 많은 영향을 미친다. 전반적인 전략은 전통적인 체계적 문헌고찰과 유사하며 숙련된 정보검색 전문가가 필요하다. 하지만 검색원과 회색문헌의 범위에서는 차이가 날 수 있다.

질문을 정할 때 검색을 시작할 수 있으며 프로토콜 단계에서 검색 전략을 개발하여야 한다. 충분히 자세히 기술하여 추후 재현, 수정할 수 있도록 하여야 한다. 검색 결과에 따라 포함기준을 변경할 수 있으며 사후 검색전략 수정이 가능하다.

신속 체계적 문헌고찰에서는 [표 20-4]와 같은 검색 전략을 수립할 수 있다.

표 20-4. 신속 체계적 문헌고찰의 검색 전략

① 2-3개로 데이터베이스 제한(Medline + Cochrane library)
② 언어: 영어, 한국어만
③ 출판 형태 : 전문만 / 국내에서 전자출판 형태로 이용 가능한 것만 / 회색문헌 배제
④ 출판연도 제한 : 가능함
⑤ 지역제한 : 가능함

20.2.5 문헌 선정

전통적인 체계적 문헌고찰과 전체적으로는 유사하지만 신속 체계적 문헌고찰에서는 제목/초록 단계에서 선정과정의 한 사람은 포함할 수 있으며, 특정 연구자는 배제만 담당하거나 다른 사람이 선정한 것을 검토만 하는 것과 같은 단순화도 가능하다. 체계적 문헌고찰 소프트웨어 혹은 온라인 프로그램을 이용하여 전문 선정 등의 작업에 적용할 수 있다.

대부분의 신속 체계적 문헌고찰에서는 기존 체계적 문헌고찰을 이용하는 경우가 많으므로, 신속 체계적 문헌고찰에서 논문 선정의 첫 단계는 질 높고 적절한 체계적 문헌고찰을 찾고 근거를 요약하는 것이다. 체계적 문헌고찰이 없으면 무작위배정 비교 임상시험, 비무작위배정 임상시험, 관찰 연구 등을 포함할 수 있다.

높은 수준의 자료를 먼저 선택하기 위해 연구유형 필터를 사용할 수 있으며, 이 시점에서 포함기준을 변경할 수 있다.

20.2.6 자료 추출

신속 체계적 문헌고찰에서는 대부분 1인은 추출하고 다른 1인은 검토한다(최소 10% 이상 검토하며 시간이 더 있으면 그 이상 검토함). 전통적인 체계적 문헌고찰의 경우 모든 자료에 대한 포괄적인 추출을 하지만 신속 체계적 문헌고찰의 경우는 관심이 있는 연구 유형과 중재결과 정보만 추출하는 것이 필요하다. 모든 연구에서 모든 효과 추정치를 추출하는 것은 권고되지 않고, 확고한 분석계획이 수립된 이후에 추출하는 것이 좋다. 사용하지도 않을 정보를 추출하여 시간을 낭비하지 말아야 한다. 이 시점에 연구의 질(비뚤림위험)을 평가할 수 있다(AMSTAR, Cochrane Risk of Bias 등).

20.2.7 자료의 합성

프로토콜 단계에서 결정한 기존 체계적 문헌고찰을 이용한 신속 체계적 문헌고찰을 하기로 하였다면 다음과 같은 형태의 신속 체계적 문헌고찰이 있다.

- 표준적인 체계적 문헌고찰(type 1)
- 체계적 문헌고찰만을 이용한 신속 체계적 문헌고찰(type 2)
- 체계적 문헌고찰과 1차 연구를 활용한 신속 체계적 문헌고찰(type 3)
- 1차 문헌만을 이용한 신속 체계적 문헌고찰(type 4)

어떤 형태로 신속 체계적 문헌고찰을 진행할지는 자료선택, 자료추출이 완료되기까지 결정하기 어려운 경우가 많다. 기존 체계적 문헌고찰을 이용한 신속 체계적 문헌고찰을 하기로 결정하였다고 하더라도 자료 합성 단계에서 표준적인 체계적 문헌고찰(type 1)이 가능하다면 그렇게 할 수 있다. 왜냐하면 이것이 가장 자료의 타당도를 높일 수 있기 때문이다. 하지만 시간, 범위, 자원 때문에 대부분 불가능하다. 다음 단계에서 type 2가 가능한지 검토한다. 1차 연구를 포함하는 것이 필요해지면 type 3, 4로 확대할 수 있다.

대부분의 체계적 문헌고찰이 사전계획을 강조하지만 신속 체계적 문헌고찰의 경우 사후계획 수정이 필요 할 수 있다.

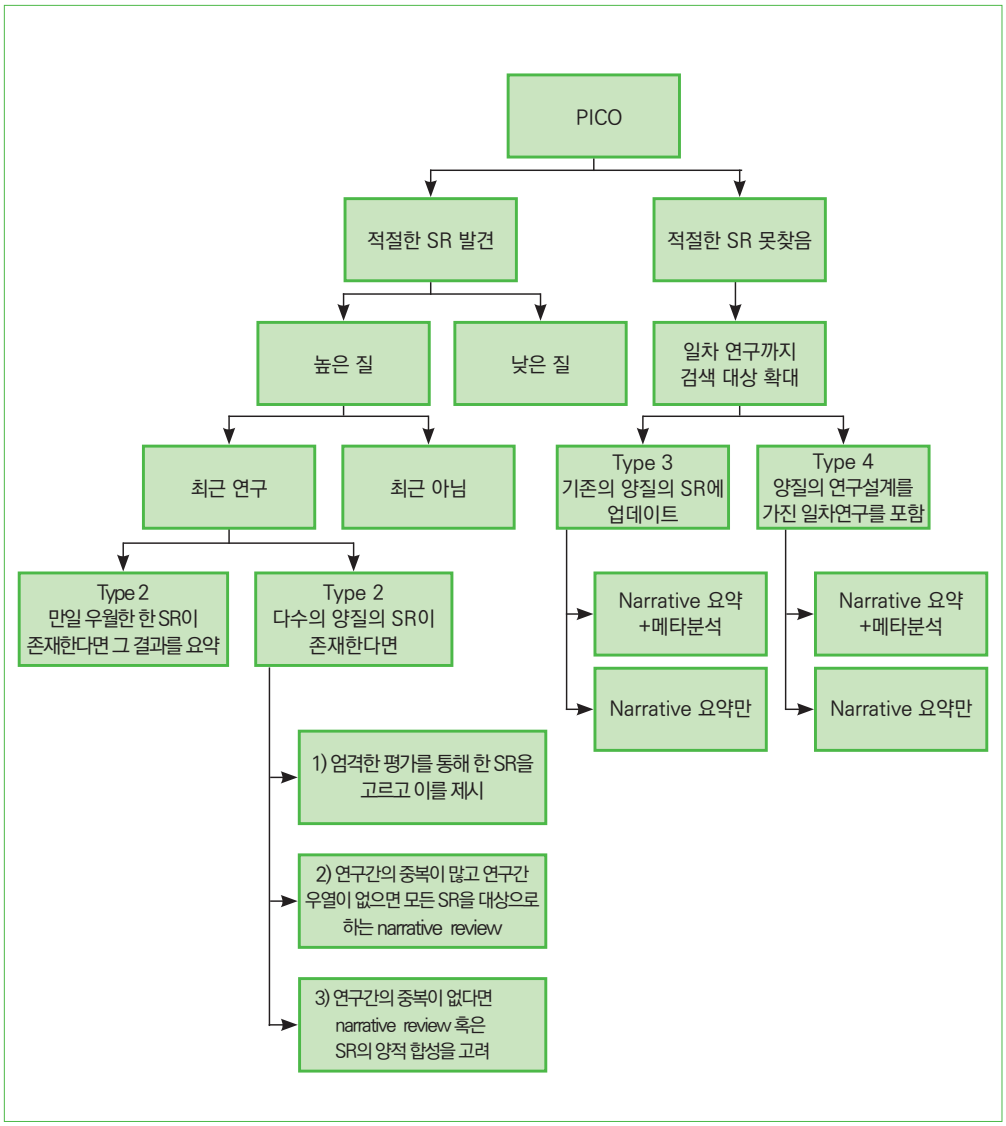


그림 20-2. 기존 체계적 문헌고찰의 이용 방법에 따른 신속 체계적 문헌고찰 방법

type 1의 시행이 어렵다면 다음과 같은 과정을 거쳐서 자료 합성을 시행한다.

첫째, 양질의 체계적 문헌고찰 여부를 확인한다.

- AMSTAR를 활용하며 8점 이상 되는 것을 높은 질의 체계적 문헌고찰로 결정할 수 있다. 참고로 AMSTAR 점수에 따른 분류는 낮음 0-3, 중등도 4-7, 높음 8-11이다.

둘째, 체계적 문헌고찰의 최신성을 평가한다.

- 이 경우 출판일보다는 검색일을 기준으로 판단한다.
- 최신성에 대한 판단 기준이 필요하다. 일부 분야의 경우 상당기간이 지나도 최신의 것으로 판단할 수 있다. 이에 대한 전문가 자문을 받는 것을 추천한다.

20.2.7.1 기존 체계적 문헌고찰의 활용(type 2)

최신의 양질의 체계적 문헌고찰이 1편이면 결과를 요약하여 그대로 사용한다. 최신의 양질의 체계적 문헌고찰이 2편 이상이면 충실도, 일차자료의 중복 정도, 연구 결론의 내용을 비교한다. 이에 따라 [표 20-5]와 같은 3가지 방법으로 진행할 수 있다.

표 20-5. 기존 체계적 문헌고찰의 활용

① 최신, 양질의 체계적 문헌고찰이 여러 편이기는 하지만 이중 한 편만을 선택하여 이것을 요약하는 것으로 결정할 수 있다.

② 포함 연구에 상당부분 중복이 있고 방법론적 충실도에서 차이가 없으면 다수의 체계적 문헌고찰에 대해 서술적 합성(narrative synthesis)를 시행한다.

③ 연구의 중복이 없고, 방법론 충실도가 유사하면 서술에 추가적으로 메타분석을 시행할 수 있다.

20.2.7.2 체계적 문헌고찰의 갱신(type 3)

이용할 기존의 체계적 문헌고찰이 존재하지만 최근 연구를 포함하고 있지 않을 때 시행할 수 있는 접근법이다. 최종 선택된 모든 체계적 문헌고찰에서 활용할 자료 목록 및 자료추출 서식을 마련한다. 일차문헌의 추가 검색은 선택한 체계적 문헌고찰들의 최종 검색일을 확인한 후, 최종검색일의 1년 전부터 현재까지 검색한다. 추가 검색된 문헌의 비뚤림위험 평가와 자료추출, 합성 등을 새로 시행할 수 있다. 결과 보고는 메타분석에 서술적 요약을 추가하여 정리하거나 서술적 요약만으로 결과를 보고한다.

자료 분석 단계에서 기존의 체계적 문헌고찰과 새로 추가된 문헌의 근거 제시는 일차연구들의 근거에 기반 해 합성되었다는 것을 분명하게 제시해야 한다. 근거를 합성할 때는 근거의 강도에 해당하는 영역들(연구의 제한점, 일관성, 정밀성, 직접성, 보고 비뚤림)을 어떻게 평가할 것인가에 대한 계획이 있어야 한다. 이때 기존 체계적 문헌고찰에 포함된 1차 연구를 다시 볼지에 대한 결정을 하여야 한다.

결과 제시를 할 때는 최종 선택된 체계적 문헌고찰을 선택한 사유와 특징을 설명하여야 한다. 양적 합성이 힘든 경우는 서술적으로 결과를 요약한다. 비뚤림위험 평가 결과(AMSTAR)를 제시하는 것이 필요하다. 필요하다면 추가 검색된 일차연구들을 추가한 경우 선택과정은 PRISMA 흐름도로 요약정리하고, 최종 문헌들의 특성을 요약 정리할 수 있다.

20.2.7.3 일차 연구에 의한 신속문헌고찰(type 4)

검색전략은 검색의 민감도와 특이도를 극대화하는 표준 검색과는 달리 특이도를 높이는 것을 지향한다. [표 5-7]과 같은 제한을 둘 수 있다.

표 20-6. 신속리뷰의 검색전략

검색원	통상 2-3개를 넘지 않음
언어	언어 제한을 흔히 사용함. 다루는 주제에 따라 적절히 결정함
접근성	전문접근이 가능한 문헌으로 국한하며, 효율을 위해 전자데이터베이스로 접근 가능한 문헌으로 한정 가능함
회색문헌	각 주제별 회색문헌의 유용성에 대해 평가하여 접근
연도	제한적 연도 검색(예, 최근 10년 등) 적용하며 이 같은 연도제한의 합당한 근거를 제시
지역	주제의 특성에 따른 지리적 위치 제한을 적용함

검색어에는 주제어(MeSH)검색과 텍스트(text)검색을 모두 포함하고, 검증된 연구설계에 대한 검색필터를 활용할 수 있다. 같은 문헌들을 대상으로 2인이 각자 독립적으로 선별하는 것과 먼저 첫 검토자가 모든 제목과 초록을 검토하고 두 번째 검토자는 첫 번째 검토에서 제외된 문헌만 검토하는 방법 중에서 채택할 수 있다.

각 핵심질문별 대상 연구 선별이 완료되면 핵심 연구별로 특성, 효과 추정값 등의 연구결과를 추출 할 수 있다. 이를 위해 표준 추출 양식을 만들고 적용한다.

보통 한 명의 검토자가 자료를 추출하고 두 번째 검토자는 추출된 모든 자료를 확인하는 방법을 취한다. 이것이 불가능하면 포함된 연구 중 적어도 10%의 무작위 표본추출을 하여 독립적으로 검사한다. 포함된 각 연구에 대해 사전에 연구계획서에 정의된 비뚤림위험평가를 수행한다. 메타 분석 같은 양적 합성은 시간과 자원의 제한으로 수행 불가능할 수 있다. 하지만 필요한 경우 시행하여 제시하도록 한다. 근거 수준은 GRADE 도구를 이용하여 시행한다.

20.2.8 보고서 작성

[표 20-7]과 같은 내용으로 보고서를 작성할 수 있다.

표 20-7. 신속 체계적 문헌고찰의 보고서 목차

개요	- 신속리뷰의 의의에 대한 간단 설명 - 수행기간 - 본 보고서는 신속리뷰이며 이런 맥락에서 해석되어야 한다는 설명
방법	- PICO 형식의 최종 핵심 질문 - 핵심(critical) 결과와 중요(important) 결과가 선택된 과정 - 연구의 포함 및 제외 기준 - 검색 전략 및 검색원 - 선별과 선택전략 - 자료추출방법 - 비뚤림위험 평가방법 - 데이터 합성과정에 대한 기술
결과	- PRISMA 흐름도를 포함한 검색 결과의 제시 - 각 핵심질문에 대한 결과요약표 -(필요하면) 각 핵심질문에 대한 GRADE 근거 평가
토론	- 신속리뷰 과정의 강점과 한계점, 특히 방법이 표준적 SR 방법과 어떻게 다른지와 신속리뷰 과정에서 발생한 비뚤림의 잠재적 위험성 - 필요한 향후 연구
추가 정보	- 감사 인사 - 저자 및 공동 연구자 목록 - 이해상충 - 재원출처 - 신속리뷰 프로세스의 한계에 관한 공개 성명
참고 자료 목록	- 포함 기준을 충족시키는 연구 목록 - 전문심사 단계에서 제외된 연구 목록 - 포함 기준을 충족할 수 있는 비영어권 또는 선택된 외국어 연구 목록 - 자료추출표 - 비뚤림위험 요약표 -(필요하면) GRADE 프로파일

보고서에는 실행요약 스타일(executive summary, 주요 메시지를 요약한 브리핑)을 첨부하며, 주요 메시지와 정책적 시사점을 앞쪽으로 배치한다. 단순함을 유지하도록 하며 간결한 문장과 단락, 표 형식 강조, 그래프 활용 등이 이루어지도록 한다. 레이아웃에 많은 공을 들이는 것이 필요하다. 하지만 어떤 형식이 더 효과가 있는지는 명확하지 않다. 어떤 매체가 더 좋은지 즉 종이, 전자매체 등 어떤 매체를 선택할지 고민이 필요하다.

20.2.9 추후관리

최종 정보사용자가 일부 내용을 추가할 수 있으며, 일단 내용에서 빠진 부분이나 근거를 잘못 해석한 부분이 없음을 최종 정보사용자가 확인하는 절차를 갖도록 한다. 최종 정보사용자와 대면회의를 통해 내용을 검토할 수도 있다. 레이아웃과 내용 모두를 검토하도록 한다. 아울

러 작업 종료 일정을 검토하여 결정하는 것이 좋다.

정보사용자(의뢰자)가 주로 참여하는 단계는 1단계: 필요성 평가, 2단계: 주제 결정, 9단계: 추후관리이다.

아래 [표 20-8]의 정보 기술은 과정의 다양한 단계를 보다 효율적으로 만들어 연구원이 신속 체계적 문헌고찰을 수행하도록 지원할 수 있다.

표 20-8. 신속 체계적 문헌고찰 수행을 위한 정보 기술(Valerie et al., 2017)

제품명	특 징
Abstrackr, OpenMeta[Analyst] (무료접근가능하며 개방된 자료원)	• Brown University School of School of Health 웹 사이트의 제품 모음으로 Abstrackr는 반자동 문헌 선별 소프트웨어 프로그램 • OpenMeta [Analyst]는 연속형, 이분형 또는 진단 검사정확도 자료의 메타 분석을 수행하기 위한 소프트웨어
Covidence (첫번째 문헌고찰시 무료; 그다음 문헌고찰부터는 구독해야 함)	• 코크란 리뷰 저자를 위한 일차 선별 및 자료추출 도구 • 비뚤림위험 평가표를 준비하기 위해 원문에 강조 표시하고 연결할 수 있음 • 다양한 분석 패키지로 자료를 내보낼 수 있음
DistillerSR (라이센스 구매 필요)	• 포함 및 배제기준을 위한 문헌 가져오기 및 추적을 위한 도구 • 사용자 맞춤형의 자료추출표 • 다양한 분석 패키지로 자료를 내보낼 수 있음
EPPI-Reviewer (코크란 저자들에게는 무료, 다른 사람들은 구독해야 함)	• 복합 문헌고찰을 포함한 모든 유형의 체계적인 문헌고찰 개발 지원 • 문헌 관리, 선별, 자료추출 및 비뚤림위험 평가 포함 • 양적 및 질적 분석 포함 • 텍스트 코딩 및 주요어(key words) 생성 가능
GRADEpro GDT (무료로 접근가능)	• 체계적 문헌고찰을 위한 근거 프로파일 및 결과 요약표를 생성하고 지침 권장 사항 개발을 지원하는 소프트웨어
Rayyan (모바일 앱을 포함한 웹 기반 소프트웨어로 무료로 접근가능)	• 반자동 문헌 제목 및 초록 선별을 위한 소프트웨어
Review Manager(RevMan) (코크란 리뷰가 아닐 경우는 라이센스 구매, 코크란 리뷰의 경우 무료로 접근가능)	• 연구 특성 표, 비교, 바이어스 위험도 평가를위한 차트 및 결과를 그래픽으로 표시하기위한 템플릿을 포함한 Cochrane 검토 템플릿 포함 • 메타 분석 소프트웨어 통합
System for the Unified Management, Assessment and Review of Information(SUMARI) (무료, 단 등록필요)	• Joanna Briggs Institute에 의해 개발된 체계적 문헌고찰을 위한 모듈 모음으로 체계적 문헌고찰 연구자들에게 접근가능 • 자료추출 및 다양한 연구설계에 대한 비평적 평가 도구를 포함함 • 서지정보를 가져오고 관리할 수 있음

20.3 참고문헌과 읽을거리

- 김수영, 신승수, 서현주, 차영주. 2017년 보건기관 만성질환 예방관리사업 대상 근거평가 및 권고개발. 2017. 한림대학교산학협력단·질병관리본부.
- Valerie J. King, Chantelle Garritty, Adrienne Stevens, Barbara Nussbaumer-Streit, Lisa Hartling, Curtis S. Harrod, Jeanne-Marie Guise, Chris Kamel. Chapter 2. Performing rapid reviews: Tricco AC, Langlois EV, Straus SE, editors. Rapid reviews to strengthen health policy and systems: a practical guide. Geneva: World Health Organization; 2017. Licence: CC BY-NC-SA 3.0 IGO.
- Tricco AC, Antony J, Zarin W, Strifler L, Ghassemi M, Ivory J, Perrier L, Hutton B, Moher D, Straus SE. A scoping review of rapid review methods. BMC Med. 2015;13:224.

제21장 경제성 근거

핵심 포인트

- 경제학은 사회에 이익을 창출하기 위해 제한된 자원의 최적 배분에 대한 연구분야이다.
- 최적의 결정을 내리기 위하여 비용 효과성에 대한 최상의 근거가 필요하다.
- 경제성 근거를 체계적 문헌고찰에 통합하는 방법에 대해 다룬다.
- 문헌고찰 질문과 근거에 대한 경제학적 관점을 문헌고찰에 통합하는 것은 보건의료 의사결정과 새로운 경제성 분석에 대한 유용성과 적용가능성을 향상시킬 수 있다.

경제학은 사회에 이익을 창출하기 위해 제한된 자원의 최적 배분에 관한 학문이다. 자원 가용성이 제한되어 있기 때문에, 의사 결정자들은 종종 중재의 이득과 유해 같은 효과성 자료뿐만 아니라 의료 자원 사용에 미치는 영향, 해당 자원 사용과 관련된 비용 및 궁극적으로 의사결정자에게 효율성에 관한 정보가 필요하다. 효과성과 효율성 자료 모두가 필요한지는 보건의료 의사결정과 밀접한 관련이 있다. 이러한 이유로, 유익하고 불리한 효과에 대한 근거와 함께, 경제적 관점과 근거를 체계적 문헌고찰에 포함시키면 문헌고찰이 의사 결정에 더 큰 도움을 줄 수 있다.

체계적 문헌고찰에 경제적 관점을 포함시킬 것인지에 대한 결정은 계획 단계에 포함되어야 한다. 경제성 근거를 체계적 문헌고찰에 통합하려면 문헌고찰에 중재의 비용이나 기타 경제적 영향의 추정치를 포함하기 위해 특수한 방법과 절차를 적용해야 한다.

이 장에서는 체계적 문헌고찰과 일치하는 환자 인구집단을 비교하는 연구에서 가져온 자원 사용 또는 비용 또는 비용 효과 자료에 관한 정보를 경제성 근거라는 용어를 사용한다. 우리가 관심을 갖는 연구 유형은 경제성 평가이다. 이는 두 개 이상의 중재의 비용과 효과를 비교하는 완전 경제성 평가(full economic evaluations) 또는 비용만 혹은 효과만을 평가하는 부분 경제성 평가(partial economic evaluations)일 수도 있다.

효과만을 비교하는 부분 경제성 평가는 이미 문헌고찰에 포함되어 있을 것이다(이 정의에서 중재의 효과와 위해를 비교하는 임상시험은 부분 경제 평가의 한 형태이다). 비용만 고려하는 부분 경제성 평가를 비용 분석이라고 하며, 근거의 일부로 포함할 수도 있다. 체계적 문헌고찰을 위한 프로토콜 계획의 초기 단계에서 각 접근법의 방법론과 실제적 의미를 신중하게 고려해야 한다. 두 가지 방법론적 틀은 다음과 같다:

- 경제성 근거를 포함하는 통합 체계적 문헌고찰(integrated full systematic review)
- 간단한 경제성 논평(brief economic commentary)

경제성 근거에 대한 통합 체계적 문헌고찰은 간단한 경제성 논평보다 실질적으로 필요한 노력의 양이 더 많다. 이는 중재 효과에 대한 주요 체계적 문헌고찰의 각 단계에 추가적인 '경제성' 방법 절차를 통합해야 하기 때문이다. 경제성 근거에 대한 종합적이고 체계적 문헌고찰을 수행하려면, 절차의 모든 단계에서 틀을 적용해본 경험이 있는 보건경제학자가 문헌고찰 팀에 투입되어야 한다.

간단한 경제성 논평 틀은 덜 집중적이고, 덜 엄격하며 이 장의 대부분은 이 접근법에 중점을 둔다. 이 틀은 특히 보건경제학자의 투입 없이 문헌고찰 팀에 큰 추가 작업 부담을 주는 범위에서 체계적 문헌고찰 틀에 경제성 근거를 포함하도록 설계되었다. 이 틀은 경제성 근거를 통합하기 위한 '최소한의 틀'로 볼 수 있으며, 논평의 한계점에 대한 적절한 주의가 필요하다.

21.1 경제성 근거 문헌고찰 방법에 대한 핵심 원칙

세 가지 핵심 원칙이 두 가지 틀 모두에 적용된다.

21.1.1 경제성 근거만 제시해서는 안된다.

중재의 비용 및/또는 비용 효과성에 대한 근거를 요약하기 위한 통합 문헌고찰 또는 간단한 경제성 논평에서 경제성 근거에 대한 연구만 단독적으로 수행하지 않아야 한다. 건강 및 관련 건강결과에 대한 중재 영향에 대해 신뢰할만한 근거의 맥락이 고려된 경제성 근거를 제시해야 한다(이 경우 자원 사용, 비용 및/또는 비용 효과성에 미치는 영향).

21.1.2 특정 자원이나 세팅보다는 경제성 중재결과(economic outcomes)에 기여한 것들을 고려한다.

체계적 문헌고찰의 최종 사용자인 전 세계의 독자를 고려할 때, 자원의 가용성 또는 보건 시스템의 구조와 같은 의사 결정 세팅에 대한 어떤 가정과 특정 자원의 추정은 적절하지 않을 수 있다. 문헌고찰의 경제성 구성 요소의 기본 목표는 특정 맥락에서 특정 시간에 구현될 때 중재가 점진적 자원 사용, 비용, 건강 결과 및 비용 효과성에 어떤 영향을 미쳤는지(즉, what happens에 초점)와 연구와 세팅 사이의 경제 및 의료 결과 추정치 변동을 유발하는 것이 무엇인지를 설명하는 것이다. 이를 통해 최종 사용자는 자신의 환경에서 실제로 사용될 수 있는 여러 대안 사이에서 주요한 경제적 저울질(economic trade-offs)를 이해할 수 있다.

22.1.3 경제성 근거가 미래의 연구에 어떻게 도움이 될 수 있는지 고려한다.

문헌고찰에서 경제성 구성 요소를 두는 주요한 목표 중의 하나는 새로운 모델 기반 경제성 평가의 형태로 진행되는 후속 혹은 병렬 연구가 필요한지를 제시하는 것이다.

21.2 문헌고찰에서 경제성 근거를 포함할지에 대한 우선순위를 정하는 기준

모든 문헌고찰이 경제성 요소를 가질 수는 있지만, 경제성 요소가 항상 필요한 것은 아니다. 일반적으로 자원 사용 및 관련 비용에 미치는 영향 측면에서, 중재군과 비교군 간에 중요한 차이가 예상될 때 경제성 근거를 문헌고찰에 포함시키는 것이 더 중요하다. 또한 전문적 지식을 가진 전문가의 가용성 및 이용 가능한 연구 자료와 같은 실용적 요소도 최종 결정에 영향을 줄 수 있다.

경제적 근거를 포함하기 위해 문헌고찰의 우선순위를 정하는 데 도움이 되는 세 가지 기준이 있다.

- 중재의 예상되는 증분 효과(즉, 대안 중재 사이의 효과 차이가 얼마나 클 것인가? 가장 작은 의미 있는 효과는 최소한의 중요한 차이 또는 환자에게 의미 있는 효과의 차이에 해당 할 수 있음)
- 중재의 예상되는 증분 비용(즉, 자원 사용의 주요 요소는 무엇이며 영향을 받을 가능성이 얼마나 큰가? 이 차이가 의사 결정자에게 얼마나 중요한가?)
- 경제성 근거가 중재의 사용에 관한 잠재적인 결정을 바꿀만한 가능성(질병의 유병률이나 보건시스템 요인과 같은 다른 맥락적 요인을 고려할 수 있음)

21.2.1 코크란 리뷰에 경제성 근거의 포함 우선순위 설정 기준 사용

표 21-1은 경제적 근거의 포함에 대한 문헌고찰의 우선순위를 정하는데 도움이 되는 결정 알고리즘이다. 예를 들어, 두 번째 행의 경우 중재의 예상 증분 효과는 작고, 증분 비용이 낮으며, 경제성 근거가 결정을 변경할 가능성이 높다. 이 시나리오는 예를 들어, 중재의 예상 비용 영향은 작지만 중재 대상이 되는 건강 상태가 작은 변화도 큰 영향을 미칠 수 있는 높은 유병률을 보일 때 발생할 수 있다. 이 경우 많은 수 환자에게 발생한 비용으로 전반적은 비용 증가가 크다고 할 수 있다.

반대로, 기대되는 증분 이득 효과가 작고, 기대되는 증분 비용이 높고, 경제성 근거가 의사

결정을 변경할 확률이 높을 경우, 이 알고리즘은 경제성 근거를 통합하는 것의 우선순위가 높다고 할 수 있다.

표 21-1. 경제적 근거의 포함에 대한 문헌고찰의 우선순위를 정하는데 도움이 되는 결정 알고리즘

기대 증분 효과	기대 증분 비용	경제성 근거가 잠재적인 의사결정을 바꿀 수 있는 확률	경제성 근거를 통합하기 위한 우선순위
작음	낮음	낮은 확률	낮은 우선순위
작음	낮음	높은 확률	중간 우선순위
큼	낮음	낮은 확률	매우 낮은 우선순위
큼	낮음	높은 확률	낮은 우선순위
작음	높음	낮은 확률	중간 우선순위
작음	높음	높은 확률	높은 우선순위
큼	높음	낮은 확률	낮은 우선순위
큼	높음	높은 확률	중간 우선순위

표 21-1에 나와 있는 의사결정 알고리즘에서 경제성 근거를 통합하는데 우선 순위가 낮거나 매우 낮은 경우, 의사 결정자에게 유용한 정보를 제공하지 않는다는 것을 반드시 의미하지는 않다. 오히려, 이는 경제성 요소를 문헌고찰에서 수행함에 있어, 제한된 연구 시간과 자원을 쏟는 데 낮은(또는 매우 낮은) 우선순위가 할당될 수 있음을 의미할 뿐이다.

예상되는 증분 이득 효과가 작고, 예상되는 증분 비용이 높고, 경제성 근거가 의사결정을 변경할 확률이 높을 경우, 이 알고리즘은 경제성 근거를 통합하는 것으로 우선순위를 두게 된다.

경제성 근거가 채택결정(adoption decisions)을 바꿀 수 있는 예상 확률은 대체로 주관적인 판단이다. 그러므로 저자는 판단을 할 때 고려해야 할 요소에 대해 전문가의 조언을 제공할 수 있는 보건경제학자에게 상담을 받을 것을 권한다.

21.3 간단한 경제성 논평 과정

21.3.1 연구목적 작성

간단한 경제성 논평의 목적은 적격한 경제성 평가의 비용 및 비용 효과 측면에서 가용성 및 주요 결과를 요약하는 것으로 주요 문헌고찰의 이차적 목적으로 기술되어야 한다.

경제적 관점을 문헌고찰에 포함시키는 목적은 문헌고찰의 대상인 건강 상태(인구집단)와 중재에 대한 '경제적 관점'을 배치하는 것이다. 이것은 문헌고찰의 배경에서 논의되어야 하며,

경제적 문제의 관련성과 문헌고찰이 다룰 질문에 대한 맥락을 강조해야 한다.

21.3.2 배경 부분에 의사결정 문제에 대한 경제적 관점 소개

문헌고찰의 배경 부분에서 강조해야 할 세 가지 경제적 이슈는 다음과 같다.

- 건강 상태의 경제적 부담(즉, '질병 비용')
- 자원 사용(비용)에 대한 중재의 잠재적 영향
- 고려해야 할 중재 비용 및 비용 효과성의 일반적인 이슈

첫 번째 이슈을 다루기 위해 배경의 '건강상태 설명' 부분을 확장하여 경제적 부담 또는 다루고자하는 건강상태의 질병 비용에 대한 논의를 포함시킬 수 있다. 이 부분의 원 자료(source material)를 알아내기 위해 간단한 문헌 검색이 필요하다.

두 번째와 세 번째 이슈는'중재가 어떻게 작동하는가' 혹은 '이 문헌고찰을 하는 것이 왜 중요한가'에 대한 배경 부분에 보고되어야 한다. 두 번째 및 세 번째 이슈의 경우 원 자료를 알아내기 위한 추가 검색이 필요하지 않다. 대신, 자원 사용에 대한 중재의 잠재적 영향과 의사결정에서의 중요성을 고려해야 한다.

검색된 질병 비용 연구의 범위에 따라'건강상태 설명' 부분의 기술에는 다음이 포함되어야 한다.

- 보건의료시스템/환자/가족/사회 전체에 대한 질병 비용/경제적 부담 규모에 대한 간략하고 일반적인 논의
- 보건의료시스템/환자/가족/사회에 대한 질병 부담을 통화로 표현하여 추정

또한 제시된 통화로 표현된 비용 추정치에는 연구가 수행된 국가, 통화 및 가격 연도(보고된 경우)에 대한 세부 정보가 포함되어야 한다.

21.3.2.1 연구배경을 위한 질병관련 비용연구 검색

배경의 '건강상태에 대한 설명' 부분에서 이러한 간단한 논의의 정보를 제공하는데 필요한 특정 유형의 보건 경제성 연구(원 자료)가 질병-비용 연구이다. 질병 비용 연구는 특정 건강 상태 또는 특정 환자 집단 내에서 사용되는 총 자원을 설명, 측정 및 평가하는 것을 목적으로 하는 경제성 평가의 한 형태이다.

이러한 검색의 목적은 다루고 있는 건강상태의 경제적 부담(질병 비용)에 대한 정보를 보고하고 있는 가장 유용한 몇 개의 논문을 찾는 것이다. 문헌을 포괄적으로 검색하고 모든 관련 연구를 검색하지 않는다. 오히려, 비용 연구가 검색될 수 있는 두 개 이상의 데이터베이스에 중점을 두어 검색하는 것이 중요하다. 가장 유용한 출처는 하나 또는 두 개의 논문이다. 이 논문에 최근 수행된 질병비용 연구 또는 최근에 수행된 개인과 사회에 대한 질병-건강의 비용 측면에서뿐만 아니라, 관리비용 측면에서 좀 더 넓은 경제적 부담 추정치와 국제적인 비교에

중점을 둔, 질병 비용연구에 대한 문헌고찰이 있으면 좋다.

질병 비용 연구를 검색하기 위해 설계된 필터와 인구집단 관련 키워드 용어는 'AND' 연산자를 사용하여 병합 검색되며, MEDLINE, EMBASE, CINAHL, PsycINFO 또는 PubMed와 같은 일반적인 전자 생의학 데이터베이스에서 수행된다. 특정 문헌고찰 주제와 관련이 있는 경우 추가 데이터베이스를 검색하여 적어도 MEDLINE 및 EMBASE 이상을 검색하는 것을 권장한다.

현재 질병 비용 연구 논문을 색인하고 있는 전문적인 보건경제학 전자 문헌 데이터베이스는 없으며, 타당성이 검증된 질병 비용 연구를 위해 특별히 설계된 검색 필터도 없다. 아래 표에 제시된 검색 필터를 사용하는 것을 권장한다.

표 21-2. MEDLINE(OvidSP) filter for cost-of-illness studies

1.(cost? adj2(illness or disease or sickness)).tw.
2.(burden? adj2(illness or disease? or condition? or economic*)).tw.
3.("quality-adjusted life years" or "quality adjusted life years" or QALY?).tw.
4. Quality-adjusted life years/
5. "cost of illness"/
6. Health expenditures/
7.(out-of-pocket adj2(payment? or expenditure? or cost? or spending or expense?)).tw.
8.(expenditure? adj3(health or direct or indirect)).tw.
9.((adjusted or quality-adjusted) adj2 year?).tw.
10. or/1-9

표 21-3. EMBASE(OvidSP) filter for cost-of-illness studies

1.(cost? adj2(illness or disease or sickness)).tw.
2.(burden? adj2(illness or disease? or condition? or economic*)).tw.
3.("quality-adjusted life years" or "quality adjusted life years" or QALY?).tw.
4. Quality-adjusted life years/
5. "cost of illness"/
6. Exp "health care cost"/
7.(out-of-pocket adj2(payment? or expenditure? or cost? or spending or expense?)).tw.
8.(expenditure? adj3(health or direct or indirect)).tw.
9.((adjusted or quality-adjusted) adj2 year?).tw.
10. or/1-9

표 21-4. PsycINFO filter for cost-of-illness studies

1.(cost? adj2(illness or disease or sickness)).tw.
2.(burden? adj2(illness or disease? or condition? or economic*)).tw.
3.("quality-adjusted life years" or "quality adjusted life years" or QALY?).tw.
4. Health Care Economics/
5. Costs and Cost Analysis/
6. Health care costs/
7.(out-of-pocket adj2(payment? or expenditure? or cost? or spending or expense?)).tw.
8.(expenditure? adj3(health or direct or indirect)).tw.
9.((adjusted or quality-adjusted) adj2 year?).tw
10. or/1-9

21.3.3 적격기준 작성

간단한 경제성 논평을 위해, 문헌고찰에 정보를 제공하기 위한 경제성 연구에 대한 인구, 중재(들), 비교 자(들) 및 결과(PICO)를 설명하는 별도의 적격기준을 포함 할 필요는 없다. 논평을 개발하기 위해 사용될 연구의 적격 기준은 PICO 요소와 관련하여 중재 효과에 대한 주요 체계적 문헌고찰을 위해 설정된 것과 동일하다.

이를 반영하기 위해 '경제성 근거 통합'이라는 방법에 섹션을 추가하여 명확하게 설명하는 것을 권한다. 이 섹션은 경제성 평가 연구설계의 유형과 관련하여 추가 기준을 기술하는 것으로 해야 한다.

표 21-5. 경제성 논평 적격기준 작성의 예시

우리는 50세 초과의 성인에서 심장마비의 일차 예방을 위해 아스피린 사용과 위약을 비교한 [임상시험과 모형에 기반한] 부분 경제성평가와 [비용분석, 비용효과분석, 비용효용분석, 비용이득분석]* 완전 경제성평가의 가용성과 주요 결과를 요약하기 위해, 최신 방법론 가이드라인(http://methods.cochrane.org/economics/)에 근거하여 간단한 경제성 논평을 개발할 것이다.

이 논평은 적격한 경제성 평가의 주요 결과가 중재가 다른 환경에서 실행될 때, 경제적 관점에서 선호되는지(또는 비선호되는지)를 판단 할 수 있는지에 초점을 맞출 것이다.

21.3.4 근거검색

문헌고찰에 포함될 연구에 대한 검색과 함께, 별도의 검색 전략을 계획해야 하며, 간단한 경

제논평 개발에 필요한 적격한 보건경제성 평가가 검색되어야 한다. 이러한 보완 검색을 위해 다음 요소가 권장된다.

- 문헌고찰에 포함된 것으로 확인된 효과 관련 연구들에서 참고문헌 목록을 점검하고 전방 인용 추적(forward citation tracking)을 수행한다.
- 중재(및 비교군) 개념에 기초한 키워드 용어를 사용하여 NHS EED(National Evaluation Database)의 검색을 수행한다.
- 중재 효과에 대한 주요 문헌고찰을 위해 검색된 하나 또는 두 개의 선택된 일반 전자 생의학 데이터베이스를 검색하여 검색된 문헌들의 세트에 전문 검색 필터를 적용한다. 관련 검색 필터의 예는 Economics Methods Group에서 얻을 수 있다.

21.3.5 문헌 선정 및 자료 수집

검색된 경제성 평가연구는 여전히 주요 중재효과에 대한 체계적 문헌고찰을 위해 이미 정의된 인구집단, 중재, 비교중재와 관련된 적격 기준에 대해 선별되어야 한다. 이 작업은 한 명의 연구자만 수행할 것을 권유한다. 이 저자는 또한 앞서 언급된 일반적인 절차를 이용하여(중재효과에 대한 주요 문헌고찰에 포함된 적격한 임상시험과의 연결을 포함하여) 경제성 평가를 분류해야 한다.

간단한 경제성 논평을 위해 자료를 수집하려면 두 가지 유형의 자료를 추출해야 한다. 즉 각각의 검색된 경제성 평가의 특성에 대한 기본 세부 사항과 주요 결과를 요약한 간단한 본문 추출이다.

각 경제성 평가의 특성에 대해 수집된 기본 자료는 다음을 포함해야 한다.

- 경제성 평가의(임상시험 또는 모델 기반) 분석 틀 및(비용 분석, 비용 효과 분석, 비용 효용 분석, 비용-이득 분석) 평가 유형(논평의 일부로 검색된 각각 유형의 수로 요약하기 위한);
- (해당되는 경우) 효과와 비용에 대해 채택된 시간대와 분석 관점(의사 결정자가 누구의 비용과 혜택을 중요하게 여기는지)과 시점(비용과 효과가 평가되는 기간);
- 각 분석에 포함 된 주요 비용 항목(예 : 보건의료 부문 비용, 기타 부문 비용, 환자 및 가족 비용 범주에 해당하는 비용 그리고 생산성이 병원 치료비용에 영향을 미치는 생산성, 직접 보건의료 비용, 간접 비보건의료 비용에 영향을 미치는 비용)에 해당하는 비용;
- 세팅(예 : 연구가 수행된 국가), 각 분석에서 사용된 통화 및 가격 연도

비용 항목을 보건 부문 비용, 기타 부문 비용, 환자 및 가족 비용 및 생산성 영향의 네 가지 범주로 분류하는 것이 도움이 된다. 보건 부문 비용에는 시스템 또는 보험자(환자가 직접 지불한 비용 제외)가 포함되며 일차의사 진료(예: 대면 방문 또는 전화 또는 인터넷을 통한 공식 진료), 처방약, 입원 및 외래 진료, 세부전문의의 3차 진료와 같은 항목이 포함될 수 있다. 기타

부문 비용에는 사회 복지 서비스, 교육 기관, 지방 당국 또는 경찰 및 사법 서비스 비용이 포함된다. 환자 및 가족 비용에는 약물 또는 돌봄에 대한 직접 지불 또는 공동 지불, 또는 교통비 또는 돌봄준비와 같은 본인부담비용을 포함할 수 있다. 생산성 손실은 경제에 대한 생산량의 손실이며, 일반적으로 이환 또는 조기사망률 뿐만 아니라 의료 서비스 이용으로 인한 근무를 못하는 시간까지 측정한다.

주요 결과를 위해 다음 자료를 수집해야한다.

- (연구 저자가 주요(기본 사례) 분석으로 보고한 내용과 관련하여) 각 경제성 평가의 저자가 도출한 결론
- 저자의 주요 결론을 둘러싼 불확실성을 요약한 본문(즉, 수행된 민감도 분석 결과를 기반으로 함)

21.3.6 근거의 비평적 평가

간단한 경제성 논평에는 포함된 경제성 평가의 방법론적 질 평가를 포함할 필요가 없다. 이 지침은 문헌고찰 팀에 부과되는 추가적인 업무량 부담을 최소화하려는 의도와 고찰에서 검색된 경제성 평가의 주요 결과에 대한 논의에 한계점을 쓰는 것 모두를 반영한다. 그러나 이 제한점은 간단한 경제성 논평의 본문에 명시적으로 기술되어야 한다. 예를 들면 다음과 같다.

> "[N]개의 검색된 경제성 평가에 대한 비평적 평가를 적용하지 않았으며, ['비교군 Y']와 비교된 ['중재 X']의 상대적 비용 또는 효율성에 관한 확고하거나 일반적인 결론을 도출하려고 시도하지 않았다.

간단한 경제성 논평의 결과는 코크란 리뷰의 고찰(결과 아님) 섹션에 통합되어야 한다. 이러한 자료를 배치하기에 가장 적합한 곳은 효과에 대한 체계적 문헌고찰 결과가 다른 정보 및 다른 고찰과 관련되어있는 부분이다.

이러한 요소의 전반적인 목표는 체계적인 근거 검색의 상세한 결과를 제시하기보다는 적합한 주의(caveats)를 통해 확인된 적격한 경제성 평가의 가용성과 주요 결과를 요약하는 것이다.

이 논평에는 다음과 같은 간략한 설명이 포함되어야 한다.

- 검색된 전자 보건경제학 문헌 데이터베이스
- 각각의 적격 비교(개별 적격 중재/비교 조합)에 대해 검색된 관련 경제성 평가의 수
- 각 연구에서 수집한 설명 정보
- (각 사례 분석과 관련하여) 각 분석의 저자에 의해 보고된 주요 결론
- (불확실성 분석 결과에 근거한) 저자의 주요 결론에 관한 주요 불확실성

간단한 경제성 논평의 요점은 간결할 수 있으며 상대적으로 그다지 많지 않은 작업의 결과에 대해 과도한 해석을 해서는 안 된다. 해석 및 논의의 요점은 확인된 주요 경제성 평가 연구 결과의 일관성에 근거하여, 중재가 경제적 관점에서 유망한 것으로 간주될 수 있는지에 대해 명확하게 판단하는 범위에 초점을 맞추어야 한다.

Box 21-1. 간단한 경제성 논평에서 간결한 논의를 위한 문장의 예

- **근거 부족**

관련 경제성 평가가 미흡한 것으로 나타났으며, ['건강 상태 Z']에 대한 ['중재 X']에 관한 경제성 근거가 현재 부족하다는 것을 나타낸다.

- **연구들 간의 모호한 결과**

['건강 상태 Z']를 가진 환자의 치료에서, ['비교군 Y']와 비교된 ['중재 X']에 대한 가용한 경제성 근거는 모호하다.

- **연구들 간 일관된 결과 [1]**

경제적인 관점에서, 이용 가능한 경제성 근거는 ['중재 X']는 ['비교 자 Y']와 비교하여 ['건강 상태 Z']의 2 차 예방을 위해 유망한 전략임을 나타낸다.

- **연구들 간 일관된 결과 [2]**

이러한 한계를 고려할 때, ['중재 X']를 제공받는 ['건강상태Z'] 환자들에 ['비교군 Y']에 비해 단기 직접 의료 비용이 평균적으로 낮다는 결과에서 경제성 평가들간에 일관성이 있었다. [' 중재 X']와 [' 비교군 Y'] 간의 주요 중재결과에 분명한 차이가 없다는 중재 효과에 관한 주요고찰에서 얻은 주요 결론과 함께 고려할 때, 이용 가능한 경제성 근거는 경제적 관점에서 ['중재 X']는 ['건강 상태 Z'] 환자에서 ['비교 Y']에 비해 안전하고 저렴한 대안으로서 유망한 중재일 수 있다.

21.4 참고문헌과 읽을거리

- Niessen L, Bridges J, Lau B, Wilson R, Sharma R, Walker D, Frick K, Bass E. Assessing the Impact of Economic Evidence on Policymakers in Health Care - A Systematic Review. Agency for Healthcare Research and Quality(US); 2012. Contract No.: No. 12(13)-EHC133-EF https://effectivehealthcare.ahrq.gov/topics/economic-evidence/research.
- Abdelhamid A, Shemilt I. Glossary of terms. In: Shemilt I MM, Vale L, Marsh K, Donaldson C editor. Evidence-based decisions and economics: health care, social welfare, education and criminal justice. Oxford: Wiley-Blackwell; 2010.

- Jenkins M. Evaluation of methodological search filters--a review. Health Information and Libraries Journal 2004; 21: 148-163.
- Drummond M, Sculpher M, Claxton K, Stoddart G, Torrance G. Methods for Economic Evaluation of Health Care Programmes. Fourth ed. USA: Oxford University Press; 2015.
- Shemilt I, Aluko P, Graybill E, Craig D, Henderson C, Drummond M, Wilson ECF, Robalino S, Vale L; on behalf of the Campbell and Cochrane Economics Methods Group. Chapter 20: Economic evidence. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA(editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.0(updated July 2019). Cochrane, 2019. Available from www.training.cochrane.org/handbook.

제22장 질적 근거 합성

핵심 포인트

- 질적 근거 합성을 통해 중재의 복합성, 맥락에 따른 변이, 실행과 이해당사자의 선호도 및 경험에 대한 이해를 높이기 위한 추가적인 근거를 제공할 수 있다.
- 이 장에서는 질적 근거 합성과정을 다음과 같이 설명 한다: 1) 질적 근거 합성 계획, 2) 질문개발, 3) 질적 근거 검색, 4) 질적 연구의 방법론적 강점과 제한점 평가, 5) 문헌선택, 6) 질적 근거 합성 방법 선택, 7) 자료추출, 8) 질적 합성 결과에 대한 신뢰도 평가, 9) 질적 근거 합성을 중재 문헌고찰에 통합하는 방법, 10) 질적 근거 합성 및 프로토콜 보고
- 질적 근거 합성은 하나의 문헌고찰 프로토콜 내에 중재문헌고찰과 함께 질적 근거 합성을 통합하는 혼합방법설계를 이용하거나 질적근거 합성을 수행하고 해당 중재 문헌고찰과 함께 통합될 수 있다.

질적 근거 합성은 중재가 어떻게 작동하는지에 대한 이해를 도울 수 있다. 관심 현상에 대한 이해를 증진시키고(예. 좋은 산전관리는 무엇인지에 대한 여성들 인식의 개념화), 세팅과 중재의 관련성을 조사하기도 한다. 또한 중재를 제공받거나 실행하는 사람들이 건강 상태나 중재에 대해 가지는 가치와 태도 및 경험에 대한 이해를 증진시키고, 중재와 실행의 복잡성, 다양한 하위군에 미치는 영향, 효과와 상황에 따른 개인과 맥락적 특성의 영향에 대한 상세한 이해를 제공한다.

혼합 방법 연구들의 질적 구성 요소나 설문 조사의 프리텍스트에서 파생된 서사자료(narrative data)처럼 다른 유형의 질적 근거 (또는 자료)가 잠재적으로 합성을 풍부하게 할 수 있기 때문에 '질적 근거 합성'(qualitative evidence synthesis, QES)이라는 용어를 사용한다. 질적 연구 영역에서는 개념적이고 맥락적인 세부사항의 수준을 기술하기 위해 특정 용어를 사용한다. 개념적 세부수준의 높거나 낮은 수준을 포함하는 질적 근거는 풍부한 또는 부족한(rich or poor)으로 기술된다. 심층적(thick) 또는 빈약한(thin)과 같은 용어는 맥락상의 세부 수준을 높이거나 낮추는데 가장 자주 언급된다.

방법론 및 방법을 선택하기 전에 사용 가능한 질적 근거의 유형과 풍부성을 결정하기 위해 주제범위 검색(scoping search)을 수행할 것을 권유한다. 중재 효과에 대한 문헌고찰에서 질적 연구를 선택할 때, 두 가지 유형의 질적 연구가 가능하다 : '형제임상시험(trial siblings)'로 알려진 포함된 임상시험과 동일한 연구대상자로부터 자료를 수집한 질적 연구와 포함된 임상시험과 관련이 없는 별도의 연구로 진행된, 중재에 대한 관련 이슈를 다룬 질적 연구이다.

둘 다 유용한 정보를 제공할 수 있으며, 형제 임상시험 연구는 포함된 임상시험에 대한 정확한 맥락이라는 관점에서 명백하게 더 가깝고, 형제임상시험이 아닌 연구는 임상시험에 존재하지 않는 관점에 기여할 수 있다.

질적 근거와 중재문헌고찰을 통합하기 위한 설계는 다음 두 가지 접근법이 있다.

1. 순차적 문헌고찰 (Sequential reviews): 1개 이상의 중재 문헌고찰이 유사한 주제로 이미 출판되어 있을 때, 혼합 방법 문헌고찰(mixed-method review)를 생산하기 위해 후속적으로 질적 근거 합성을 수행해서 그 결과를 중재 문헌고찰의 결과들과 통합함
2. 수렴 혼합방법 문헌고찰 (Convergent mixed-methods review): 기존에 중재 문헌고찰이 출판되지 않은 경우, 임상시험연구와 질적 근거를 개별적으로 합성하여 개발과정에서 서로 '이야기'를 할 기회를 만들고, 세 번째 합성과정에서 통합시키는 완전 수렴 '혼합 방법' 문헌고찰 (full convergent 'mixed-methods' review)을 수행할 수 있음

최소한 한 명의 공통 저자가 두 문헌고찰을 모두 참여하여 질적 근거 합성 프로토콜의 개발, 합성 수행 및 혼합 방식 문헌고찰 내에서의 중재문헌고찰과 질적 근거 합성의 후속적인 통합을 용이하게 할 것을 권유한다.

22.1 중재고찰과 연계된 질적 근거 합성 계획

The Cochrane Qualitative and Implementation Methods Group (QIMG) website에는 QES를 고려하는 저자들에게 실용적인 지침과 주요 단계들에 대한 링크를 제공하고 있다 (methods .cochrane.org/qi). The RETREAT framework은 문헌고찰을 계획할 때 문헌고찰 저자가 체계적으로 수행해야하는 7가지 주요 고려사항에 대한 것이다(Booth et al 2016, Booth et al 2018) (Box 22-1).

Box 22-1. 질적 근거 합성을 위한 적절한 방법을 선택하기 위한 RETREAT 고려사항

- Review question(문헌고찰 질문) - 먼저 문헌고찰 질문의 복합성을 고려한다. 어떤 요소가 복합성)에 가장 큰 영향을 주는가(예 : 건강상태, 중재 또는 맥락(상황)? 어떤 요소가 관심을 두어야 하는 우선순위가 높은 요소인가?
- Epistemology(인식론) - 주요 연구의 철학적 기반을 고려한다. 인식론적 고려사항에 덜 의존하는 주제합성(thematic synthesis)과 같은 방법을 선호하는 것이 적절한가?
- Timeframe(시간일정) - 이용 가능한 기간 내에 어떤 유형의 질적 근거 합성이 실행가능하고 관리가능한지 고려한다.
- Resources(자원) - 문헌고찰에 대한 야망(ambition)이 사용 가능한 자원과 일치하는지 고려한다. 문헌고찰의 범위와 추출 접근법의 범위가 제한될 필요가 있는가?
- Expertise(전문성) - 문헌고찰 팀 내에서 그리고 자문그룹에서 전문지식 보유를 고려한다. 이용가능한 전문성이 선택된 질적 근거 합성 접근법에 부합하는가?
- Audience and Purpose(청중과 목적) - 의도된 청중 및 문헌고찰의 목적을 고려한다. 질문 수립, 문헌고찰의 범위, 의도된 중재결과는 그들의 필요를 충족시키는가?
- Type of data (자료유형) - 포함된 연구에 존재하는 자료의 유형을 고려한다. 고려하고 있는 연구는 개념적으로 풍부하고 맥락상 어느 정도까지 심층적인가?

22.2 질문 개발

일부 질적 문헌고찰은 양적합성의 PICO 구조를 채택하거나 PICO 구조를 변형하여 사용한다. (예. SPICE : Setting, Perspective, Intervention or Phenomenon of Interest, Comparison, Evaluation; SPIDER : Sample, Phenomenon of Interest, Design, Evaluation, Research type) (Cooke et al 2012). 질적 근거 합성과 복합중재 문헌고찰에 특히 적합한 광범위한 맥락과 즉각적인 세팅을 모두 설명하는 extended question framework (PerSPecTIF)도 있다(Booth et al 2019). 시스템 관점이 적용된 PerSPEcTiF framework 사용을 통해 실행 가능성 및 수용 가능성과 같은 보다 적절한 문제를 다룰 수 있게 된다(표 22-1).

중재가 어떻게 작동하는지, 그리고 실제로 어떻게 실행되는지와 같은 중재실행에 대한 이해를 통해 관찰된 효과성 부족의 원인이 개념적으로 빈약한 중재(즉, 이론 실패) 때문인지 빈약한 중재 실행 (즉, 실행 실패) 때문인지 검토할 수 있다.

이질성이 있다면 기본 이론과 중재 실행 방식 모두를 고려해야 한다. 사전 scoping review, 개념분석(concept analysis), 비평적 문헌고찰(critical review) 또는 텍스트 내러티브 합성(textual narrative synthesis) 등을 통해 중재나 프로그램 이론, 로직 모델 또는 실행 측정 및

과정을 조사할 수 있다. The intervention Complexity Assessment Tool for Systematic Reviews iCAT_SR (Lewin et al 2017)은 중재의 복합성을 분류하고 관련 질문을 개발하는데 도움이 될 수 있다. 명백한 프로그램 이론과 실행이 중재결과와 어떻게 관련되는지에 대한 세부 사항이 없으면 선험적 실재주의고찰(a priori realist review), 메타-민속지학(meta-ethnography)나 메타 해석적 문헌고찰(meta-interpretive review)이 수행될 수도 있 (Booth et al 2016)

문헌고찰 저자는 건강 문제 및 중재의 영향을 받는 소비자, 이해 관계자의 참여를 고려하여야 하며 이들이 임상적 또는 정책적 상황의 고려를 할 수 있게 해야 한다. 준비 단계에서 이러한 자문을 통해 문헌고찰의 범위와 질문을 적절히 하고, 문헌고찰 결과가 의사 결정자의 실행 문제를 해결할 수 있게 할 수 있다.

표 22-1. 질적 근거 합성을 위한 PerSPecTIF 질문형식 틀 (Booth et al (2019b). Reproduced with permission of BMJ Publishing Group

Per	S	P	E	(C)	Ti	F
Perspective	Setting	Phenomenon of interest/ Problem	Environment	Comparison (optional)	Time/ Timing	Findings
임산부의 관점에서	농촌지역사회 세팅에서	의료기관 기반 서비스는 어떠한가?	멀리 위치한 의료기관과 운송수단이 열악한 환경에서	가정에서 전통산파와 비교하여	분만 때까지	여성의 인식과 경험과 관련하여

22.3 질적 근거 검색

양적 연구의 검색(제 4 장)과 비교할 때, 질적 연구의 검색 절차는 상대적으로 덜 표준화 되어 있다. 질적 근거 검색에서 특히 극복해 할 과제는 제목과 초록에 정보가 없는 것(non-informative titles), 산만한 용어, 빈약한 색인, 부족한 논문 수 등이 있다(Booth et al 2016). 질적 연구 검색을 계획 할 때 주요 고려 사항을 7S framework(Sampling, Sources through Structured questions, Search procedures, Strategies and filters and Supplementary strategies to Standards for Reporting)을 사용한다(Booth et al 2016).

Box 22-2. 7S framework

Sampling : 질적 근거 합성의 목적에 부합하는 핵심 결정은 정량 문헌고찰 검색에 주로 이용되는 포괄적이고 철저한 접근법을 사용할지 아니면 질적 패러다임에 보다 민감한 목적적 샘플링을 사용할지 여부이다. 목적적 표본추출은 이론을 생성할 때 그리고 해석적 이해를 생성하기 위해 사용되며, 이론 표본추출 (theoretical sampling), 최대 변동 표본추출 (maximum variation sampling) 및 강도 표본추출 (intensity sampling) 등이 있다.

Sources through Structured questions: 질적 근거의 출처는 표준적인 양적 연구 보고서보다 단행본, 회색 문헌 보고서를 포함할 가능성이 높으므로 검색전략은 이러한 출처에 중점을 두어야 한다. 지역 데이터베이스 (Local databases)는 맥락의 중요성을 감안할 때 특히 중요할 수 있다.

Strategies and filters and Supplementary strategies to Standards for Reporting : 연구 필터(filters)를 사용할지 아니면 단순히 연구 선정 단계에서 질적 연구가 확인되는지 주제 기반 검색(topic-based search)을 수행할지 여부이다. 질적 연구를 위한 검색 필터는 양적 연구의 검색필터에 비해 특이성이 부족하다. 그럼에도 불구하고 특히 양적 문헌이 압도적으로 많아 검색되어 검색 필요 수(number needed to search)가 증가하는 경우, 필터가 필요하며, 연구유형, 혼합방법론, 관점 등을 적용할 수 있다.

과정평가와 실행 근거에 대해 검색하는 데는 몇 가지 접근법이 있다. 아래의 네 가지 접근법이 가능하다.

1. 맞춤형 검색 전략을 통해서보다는 연구 선정 시점에서 연구를 식별한다. 여기에는 연구유형 필터없이 민감한 주제 검색을 수행하고 선별 과정에서 관심 있는 모든 연구 유형을 확인하는 작업이 포함된다. 이 접근법은 문헌고찰 질문에 여러 출판 유형(예. 무작위임상시험, 질적연구, 경제성평가)이 포함 되는 경우에 유용하다.
2. 표준 검색 필터를 사용하여 무작위임상시험 내에 수행된 과정평가 연구만으로 제한한다.
3. 과정 평가 또는 실행 자료를 검색하기 위해 평가되지 않은 필터 용어(예 : 'process evaluation', 'program(me) evaluation', 'feasibility study', 'implementation' or 'proof of concept'등)을 사용한다. 연구 유형 또는 목적과 관련된 용어의 문자열을 사용하는 접근법은 실험적으로 간주되며, 이러한 필터는 다시 개발되고 검증될 필요가 있다.
4. 주제 기반 검색(topic-based searching)에 대한 의존도를 최소화하고, 인용기반 접근법(citations-based approaches) 을 사용한다. 이를 통해 실행이나 과정 자료를 제공해줄 수 있는 특정 연구의 링크된 보고서, 출판되거나 출판되지 않은 연구를 검색한다.

22.4 질적 연구의 방법론적 강점과 제한점 평가

질적 연구에 대해 방법론적 강점과 제한점에 대한 평가하는 것은 질적 연구그룹 내부에서도 여전히 논쟁의 여지가 있다. 그러나 체계적인 문헌고찰 및 근거 합성에서 방법론적 강점과 제한점에 대한 평가는 필수적이다.

연구의 엄격성의 지표로서 연구의 질에 대한 평가(assessment of study quality)를 하는 것과 비뚤림위험평가는 구별할 필요가 있다. 후자는 연구의 업격성을 평가하기 위해 방법론적 강점과 제한점을 평가하는 것이다(엄격성 위험 평가, 'risk to rigour' approach). 하지만 명백한 엄격성 평가도구가 없기 때문에 대안으로 방법론적 강점과 질적연구의 한계점을 평가하는데 초점을 맞춘 타당도가 검증된 도구를 선택할 것을 권유한다(Box 22-3).

Box 22-3. 연구엄격성 평가를 위한 방법론적 강점과 제한점 평가 영역

- 명확한 목표와 연구 질문
- 연구 목적 / 질문과 연구설계 / 방법 간의 조화
- 질문을 해결하기 위한 사례나 연구대상자 확인, 표본추출, 자료수집의 엄격성
- 방법의 적절한 적용
- 결과의 풍부함 / 개념적 깊이
- 연구자의 반추(Reflexivity)*

*반추는 질적 연구자와 문헌고찰 저자가 맥락(상황), 연구대상자, 자료 및 결과의 해석 및 보고에 대한 연구자의 실제적이고 잠재적인 영향을 고려하는 것이다. 반추하는 것은 이해 상충을 투명하게 밝히고, 문헌고찰 과정 및 결과에 대한 문헌고찰의 영향을 논의하고, 토론된 모든 문제와 후속 결정을 투명하게 만들어주는 것을 의미한다.

엄격성 위험에 대한 보완적 또는 상반되는 관점의 확인이 합의에 도달하는 그 자체보다 더 유용하다고 간주되는 경우 평가자 간 신뢰도와 같은 문제는 덜 중요하다(Noyes et al 2019).

다른 비뚤림위험평가 도구와 마찬가지로 도메인에 점수를 적용하거나 전체 질 평가 점수를 계산하지 않는 것을 권유한다. 문헌고찰 저자들이 각 논문의 '엄격성 위험 평가도구'를 평가하고, 연구의 방법론적 제한이 문헌고찰 결과에 어떤 영향을 미칠 수 있는지에 대해 토론할 것을 권장한다. 합성의 견고성과 방법론적으로 제한된 연구에 대한 취약성을 탐구하는 질적 '민감도 분석' 이 합성된 결과에 대한 문헌고찰 저자에 대한 전반적인 신뢰도에 관계없이 일상적으로 적용될 것을 권고한다. 근거는 질적 요소가 추출된 혼합 방법 연구에서 질적 민감도 분석이 똑같이 권장된다는 것을 제시한다.

몇 가지 평가 도구는 과정평가 또는 실행 근거의 엄격함을 명시적으로 다루고 있다. 일차 질적 연구의 경우, EPPI 센터에 의해 개발 된 8 개 항목 과정 평가 도구를 사용하여 초기 질적 연구에서 엄격한 방법론적 강점과 제한점 및 위험을 평가하기 위해 선택된 도구를 보완할 수 있다(Shepherd et al., 2010). 이러한 항목 중 하나인 유용성에 대한 질문, 중재 과정은 얼마나 잘 기술되었는지, 그리고 과정 자료는 중재가 왜 또는 어떻게 작동했는지 또는 작동하지 않았는지를 제시해주고 있는지 등')은 과정 기전을 탐색하는 기전을 제공한다.

22.5 문헌 선택

연구의 포함 또는 배제에 관한 결정은 모든 관련 연구를 포함하려는 무작위임상시험에 대한 문헌고찰과 비교할 때, 질적 근거 합성에서 더 복잡할 수 있다. 모든 연구를 포함할지 또는 연구들의 표본을 선택할지에 대한 결정은 일반적인 그리고 문헌고찰의 특이적인 기준에 달려있다 (Noyes et al., 2019). 선택된 질적 연구의 수는 관리 가능한 합성(manageable synthesis)과 일치해야 하며, 포함된 연구의 맥락은 효과성 분석에서 임상시험과 통합 가능해야한다. 원칙적으로 모든 결정 및 그 이유는 투명해야 한다.

22.6 질적 근거 합성법 선택

일반적으로 이용 가능한 질적 근거의 풀(pool)이 완전히 파악이 될 때까지 적절한 합성 방법을 선택하기 힘들다. 유동적인 합성 방법 중 선택을 어떻게 할 것인가에 대한 내용이 프로토콜에 명시되어야 할 수도 있다. 질적 근거 합성과 의료기술평가를 위한 방법론 선택에 관한 INTEGRATE-HTA 지침은 합성 방법을 선택할 때 유용한 출발점을 제공한다 (Booth et al 2016, Booth et al 2018). 일부 방법은 주로 기술적인 수준(descriptive level)에서 결과를 합성하도록 해서 정책 및 임상진료에 곧바로 적용할 수 있도록 설계되었다.

새로운 이론을 개발할 수 있도록 하는 방법론도 있다 (예를 들어, 메타 민속지학(meta-ethnography) 와 이론 합성 (thematic synthesis) 에 대한 이론적 접근 방법). 세 가지 질적 근거 합성 방법 (이론 합성, 틀 합성 및 메타-민속지학)은 순차적 중재 문헌고찰이나 분석에 통합될 수 있는 합성 결과물을 생산하기 위해 권장 된다(표 22-2).

표 22-2. 중재문헌고찰이나 혼합방법 문헌고찰로 통합하기 위한 질적 근거 합성 수행시 권고되는 방법론

방법론	설명
가장 적합할 가능성이 있음	
이론적 합성(Thematic synthesis) (Thomas and Harden 2008)	• 장점 : 가장 접근하기 쉬운 형태의 합성. 명확한 접근 방식으로 “빈약한” 자료로 기술적인 이론을 생성하고, 좀 더 심층적인 분석 이론들에서, 기술적인 이론을 개발하기 위해 “좀 더 심층적인”자료를 사용할 수 있다. 그런 다음 이론들은 정량적 합성 내에 통합된다. • 단점 : 해석적 '능력'에서 제한적일 수 있으며, 지나치게 단순한 사용으로 인해 가이드라인과 같은 의사 결정에 영향을 주지 않을 수 있다. 숙련된 팀이 필요한 복잡한 합성 과정. 이론적 결과들은 가설을 형성하기 위해 경험적 증거, 전문가 의견 및 추측을 결합할 수 있다. GRADE CERQual이 합성된 질적 결과에 대한 신뢰도를 평가하기 위해 이론적 결과에 어떻게 적용되는지에 대한 더 많은 연구가 필요하다. 더 높은 수준의 결과가 실행 가능한 지점으로 변환되는 방법에 대한 명확성이 부족할 수 있다.

사용시 약간 주의가 필요함	
틀 합성(Framework synthesis) (Oliver et al 2008, Dixon-Woods 2011)	• 장점 : 이론 표현을 포함하여 틀 내에서 복합성을 수용함으로써 복합 중재에 대한 문헌고찰 내에서 잘 작동한다. 이 틀은 질적 근거와 양적근거를 종합적으로 통합할 수 있는 명확한 기전을 제공한다. 중재의 본질과 중재가 의도하는 영향에 대한 광범위한 합의가 있는 경우 잘 작동한다.
최적합 틀 합성(Best-fit framework synthesis) (Carroll et al 2011)	• 단점 : 틀을 식별하고, 선택하며 정당화가 필요하다. 추출/ 합성이 진행된 이후에만 틀이 부적절한 것으로 밝혀 질 수 있다. 편의를 위해 자료를 틀에 단순화시키는 위험이 있다.
사용시 더 많은 주의가 필요함	
메타 민속지학(Meta-ethnography) (Noblit and Hare 1988)	• 장점 : 주로 해석적 합성 방법을 사용하여 새로운 고차 구성(contructs) 뿐만 아니라 기술적 결과물을 가져온다. 설명적이고 이론적인 결과는 가이드라인과 같은 의사 결정에 도움이 될 수 있다. 명시적 보고 표준이 개발되어 있다. • 단점 : 숙련된 팀이 필요한 복잡한 방법론 및 합성 과정이다. 다른 방법론보다 많은 시간과 자원이 소요될 수 있다. 이론적 결과들은 가설을 형성하기 위해 경험적 증거, 전문가 의견 및 추측을 결합할 수 있다. 감사 추적에 대한 요구 사항을 충족하지 못할 수 있다 (새로운 보고 지침이 이를 극복하는 데 도움이 되지만 (France et al. 2019). CERQual이 이론적 결과들에 어떻게 적용될 수 있는지 결정하려면 더 많은 작업이 필요하다. 더 높은 수준의 결과가 어떻게 실행 가능한 포인트로 전환되는지 확실하지 않을 수 있다.

22.7 자료추출

질적 연구 결과는 연구대상자로 부터의 발췌문, 연구저자에 의해 확인된 하위주제와 주제, 설명(explanations), 가설 또는 새로운 이론 등의 형태를 취한다. 혹은 내러티브(서사)로 제시되거나 표, 그림 또는 로직 모델로 요약되고 표시될 수 있다. 질적 자료추출 방법은 선택한 합성 방법에 따라 다르다. 자료추출은 순차적거나 선형적이지 않다. 종종 검토 단계마다 앞뒤로 이동해야 할 수도 있다. 문헌고찰 팀은 근거를 논의하고 질문하고 공유된 이해에 도달하기 위해 정기적인 회의가 필요하다. 주요 이해 관계자 그룹은 근거를 해석하고 주요 결과를 도출하는 데 도움이 될 수 있다.

문헌고찰 유형과 합성 방법 선택과 관계없이, 각 연구에 대한 상세한 맥락 및 방법론 정보를 추출하고 이 정보를 '포함된 연구의 특성' 표에 보고하는 것을 권유 한다(표 22-3). 중재 기술에 대한 템플릿 TIDieR 체크리스트(Hoffmann et al 2014)와 ICAT_SR 도구는 추출을 위한 주요 정보를 구체화하는데 도움이 될 수 있다(Lewin et al 2017). 문헌고찰 저자는 일차 연구의 잘못된 이해를 막기 위해 추출 및 합성 과정에서 일차 연구 자료 내용을 보존해야 한다.

표 22-3. '포함된 연구의 특성' 표에 포함될 맥락 및 방법론 정보

자료추출 분야	추출된 정보
맥락 및 연구대상자	문헌고찰 질문을 다루고 있는 일차 연구 맥락의 주요 요소들 (예. 연구 세팅, 인구집단 특성, 연구대상자 특성, 전달된 중재 (해당되는 경우) 등)
연구설계 및 사용된 방법	연구에서 사용된 방법론적 설계와 접근법; 표본을 모집하기 위한 방법; 구체적인 자료수집과 사용된 분석법, 연구결과를 맥락화하거나 해석하기 위해 사용된 이론적 모형들

Noyes et al (2019)은 자료 추출의 다양한 방법에 따른 추가적인 지침과 예를 제공한다. 문헌고찰 저자는 한 가지 방법을 선택하는 것이 일반적이다. 요약하면 추출 방법은 다음과 같이 그룹화 할 수 있다.

• 일반적으로 맞춤형, 표준화되거나 개조한 자료추출 형식 사용

문헌고찰 저자는 자체 문헌고찰 특이적인 자료추출형식을 개발하거나 연구 유형별로 일반적인 자료추출 형식을 선택할 수 있다 (예 : the National Institute for Health and Clinical Excellence (National Institute for Health Care Excellence 2012)에서 개발한 템플릿).

• 선험적 이론 또는 사전에 정의된 틀을 사용하여 자료 추출

틀 합성(Framework synthesis) 및 하위 변형인 '최적 맞춤' 틀 (Best Fit' Framework) 접근법은 관심 있는 현상을 더 잘 이해하기 위해 선험적 기본 틀을 이용하여 일차연구 자료를 추출하는 것과 관련이 있다. SURE 틀은 보건의료시스템 중재의 실행에 영향을 미칠 수 있는 가능한 요소들을 열거한다. (The SURE (Supporting the Use of Research Evidence) Collaboration 2011). 'PROGRESS' (place of residence (거주지역), race/ethnicity/culture/language (민족/인종/문화/언어), occupation(직업), gender/sex(성별), religion(종교), education(교육), socioeconomic status(사회경제적 지위), and social capital(사회적 자본)) 틀의 사용은 또한 명백하게 형평성에 중점을 두어 자료추출을 지속하는 것을 돕는다. 로직모델 또한 자료추출을 위한 틀로서 사용될 수 있다.

• 일차연구를 귀납적으로 코딩하기 위하여 소프트웨어 프로그램 이용

일차 질적연구 자료를 분석하기 위해 NVivo[4]와 같은 소프트웨어를 사용할 수 있다. 예를 들어, 자료추출 및 이론 합성(thematic synthesis)을 위해 관심현상에 대한 귀납적인 기술적 이론(descriptive theme)과 이론적 설명을 소프트웨어 프로그램을 이용하여 코딩하는 것이다.

4) www.qsrinternational.com/nvivo/home

22.8 질적 합성 결과에 근거 수준 평가

GRADE 시스템은 오랫동안 정량적 결과 근거 수준 평가를 위해 이용되었고, 이와 대응되는 GRADE-CERQual의 적용은 Cochrane 질적 근거 합성에서 권장된다(Lewin et al 2015). CERQual에는 합성된 질적 결과에 대한 전반적인 근거 수준을 공식화하는데 사용되는 네 가지 구성 요소(관련성, 방법론적 제한, 적합성 및 일관성)가 있으며, 구성 요소 및 보고지침 은 Implementation Science(Lewin et al 2018)의 시리즈를 확인하도록 한다.

22.9 질적 근거 합성을 중재문헌고찰에 통합하는 방법

다양한 방법과 도구를 통해 혼합방법 합성, 자료 합성이 가능하다(Harden et al 2018, Noyes et al 2019). 문헌고찰 저자는 유사한 주제(순차적 접근법) 대해 출판된 기존의 중재 문헌고찰과 질적 근거 합성을 통합하거나, 통합 전에 새로운 중재문헌고찰 및 질적 근거 합성을 병행할 수 있다(수렴 접근법).

질적 합성이 순차적인지 또는 중재문헌고찰에 수렴하는지에 관계없이, 질적 및 양적근거는 통합 전에 적절한 방법을 사용하여 개별적으로 합성되는 것이 바람직하다(Harden et al 2018). 문헌고찰 저자가 중재 문헌고찰 보고서의 기초가 되는 자료에 접근할 수 있지 않는 한 통합 범위는 기존의 중재문헌고찰로 제한할 수밖에 없다.

Harden et al과 Noyes 등은 중재문헌고찰과 통합하기 위해 다음의 방법과 도구을 설명하였다(Harden et al 2018, Noyes et al 2019):

- 결과를 매트릭스에 병치

병치는(juxtaposing) 질적 근거 합성(예 : 중재의 수용 가능성 또는 실행 가능성과 관련된 중재 요소)에서 얻은 결과에 의해 주도되며, 이러한 결과들은 매트릭스의 한 측면을 형성한다.

중재 효과에 대한 결과(예 : 결과 개선, 결과 차이 없음, 불확실한 효과)가 매트릭스의 다른 면을 형성한다. 양적 연구는 중재효과에 대한 결과들과 질적 합성에서 생성된 가설에 의해 구체화된 특징의 유무에 따라 분류된다. 행렬에서 관측된 패턴은 양적연구 결과의 차이점을 설명하고 연구에서 간극을 확인하는데 사용된다.

- 프로그램 이론 분석

중재 작동에 대한 설명 이론들이 분석된다. 중재 효과를 검증하는 양적연구, 질적 연구 및 과정 평가 근거 결과들을 이론이 실제로 어떻게 작동하는지를 조사하기 위해 사용한다(Greenhalgh et al 2007). 다양한 이론의 가치가 평가되거나 새로운/개정된 이론이 개발된다. 중재 효과성을 높이거나 낮추는 요인들도 제시된다.

• 로직모형이나 다른 유형의 개념적 틀 사용

중재가 변화를 일으키는 과정을 나타내는 로직 모델 또는 다른 유형의 개념적 틀은 다양한 유형의 근거 결과를 통합하기 위한 공통의 골격을 제공한다(Booth and Carroll 2015). 틀은 문헌이나 이해 관계자 참여를 통해 구체화되거나 문헌고찰 중에 새로이 개발될 수도 있다. 중재의 효과를 검증하는 양적 연구와 질적 근거로부터의 결과들은 모델을 개발하거나 정교화 하는데 사용된다.

• 질적 근거 합성에서 파생된 가설 검증

양적연구는 검증될 가설에 명시된 명제의 유무에 따라 분류되며 하위 집단 분석은 중재의 영향에 대한 차이점을 조사하는 데 사용된다.

• 질적 비교 분석

질적 합성에서 얻은 결과는 성공적인 중재를 위한 중요한 특징의 범위와 이러한 특징이 작용하는 기전을 확인하는 데 사용된다. 질적 비교 분석은 그 특징들이 효과적인 중재와 관련이 있는지 여부를 검증한다. 분석은 맥락에 따라 동일한 중재 특징이 효과적이고 덜 효과적인 중재와 관련된 시나리오를 조정하는 효과성에 대한 여러 잠재경로를 풀어낸다.

질적 비교 분석은 통합에 사용할 수 있는 잠재력을 제공한다. 여기에 제시된 다른 방법 및 도구와 달리 아직 방법론적 지침이 충분하지 않다. 그러나 질적 비교 분석을 사용한 모범적인 문헌고찰이 접근가능하다.

문헌고찰 저자는 위에서 제시한 방법들을 조합하여 사용할 수 있다(예. 매트릭스 내의 결과들을 병치하여 관찰된 패턴은 하위집단 분석 또는 질적 비교 분석을 사용하여 검증할 수 있다). 로직모델과 질적 비교 분석을 사용하여 프로그램 이론을 분석하기 위해서는 문헌고찰 팀 구성원이 특정 기술을 소유하고 있어야 한다. 하위 집단 분석과 질적 비교 분석을 사용하는 것은 근거가 제한적으로 존재할 때 적합하지 않다.

22.10 질적 근거 합성 및 프로토콜 보고

Reporting standards and tools designed for intervention reviews(such as Cochrane's MECIR standards(http://methods.cochrane.org/mecir)나 PRISMA Statement는 질적 근거 합성이나 통합된 혼합방법 문헌고찰에 적절치 않다. 문헌고찰 저자는 다음과 같은 사항을 고려해야한다. 문헌고찰 저자는 특정한 보고지침이 이용가능한지(예 : 메타 민속지학을 위한 eMERGe(France et al 2015)), 일반적인 지침이 적합한지(예 : ENTREQ(Tong et al 2012)), 또는 추가 점검표 또는 도구가 문헌고찰의 특정 측면을 보고하기 적합한지에 대해서 고려해야 한다.

22.11 참고문헌과 읽을거리

- Jane Noyes, Andrew Booth, Margaret Cargo, Kate Flemming, Angela Harden, Janet Harris, Ruth Garside, Karin Hannes, Tomás Pantoja, James Thomas. Chapter 21: Qualitative evidence: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA(editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.0. Cochrane, 2019. Available from https://training.cochrane.org/handbook/current/chapter-21
- Hong QN, Pluye P, Bujold M, Wassef M. Convergent and sequential synthesis designs: implications for conducting and reporting systematic reviews of qualitative and quantitative evidence. Syst Rev. 2017;6(1):61.
- Harris JL, Booth A, Cargo M, Hannes K, Harden A, Flemming K, Garside R, Pantoja T, Thomas J, Noyes J. Cochrane Qualitative and Implementation Methods Group guidance series-paper 2: methods for question formulation, searching, and protocol development for qualitative evidence synthesis. J Clin Epidemiol. 2018;97:39-48.
- Noyes J, Booth A, Flemming K, Garside R, Harden A, Lewin S, Pantoja T, Hannes K, Cargo M, Thomas J. Cochrane Qualitative and Implementation Methods Group guidance series-paper 3: methods for assessing methodological limitations, data extraction and synthesis, and confidence in synthesized qualitative findings. J Clin Epidemiol. 2018;97:49-58.
- Harden A, Thomas J, Cargo M, Harris J, Pantoja T, Flemming K, Booth A, Garside R, Hannes K, Noyes J. Cochrane Qualitative and Implementation Methods Group guidance series-paper 5: methods for integrating qualitative and implementation evidence within intervention effectiveness reviews. J Clin Epidemiol. 2018;97:70-78.
- Flemming K, Booth A, Hannes K, Cargo M, Noyes J. Cochrane Qualitative and Implementation Methods Group guidance series-paper 6: reporting guidelines for qualitative, implementation, and process evaluation evidence syntheses. J Clin Epidemiol. 2018;97:79-85.
- Booth A, Noyes J, Flemming K, Gerhardus A, Wahlster P, van der Wilt GJ, Mozygemba K, Refolo P, Sacchini D, Tummers M, Rehfuess E. Guidance on choosing qualitative evidence synthesis methods for use in health technology assessment for complex interventions 2016. https://www.integrate-hta.eu/wp-content/uploads/2016/02/Guidance-on-choosing-qualitative-evidence-synthesis-methods-for-use-in-HTA-of-complex-interventions.pdf
- Booth A, Noyes J, Flemming K, Gehardus A, Wahlster P, Jan van der Wilt G, Mozygemba K, Refolo P, Sacchini D, Tummers M, Rehfuess E. Structured methodology review identified seven(RETREAT) criteria for selecting qualitative evidence synthesis

approaches. Journal of Clinical Epidemiology 2018; 99: 41-52.
- Booth A, Noyes J, Flemming K, Moore G, Tuncalp Ö, Shakibazadeh E. Formulating questions to address the acceptability and feasibility of complex interventions in qualitative evidence synthesis. BMJ Global Health 2019; 4: e001107.
- Lewin S, Hendry M, Chandler J, Oxman AD, Michie S, Shepperd S, Reeves BC, Tugwell P, Hannes K, Rehfuess EA, Welch V, McKenzie JE, Burford B, Petkovic J, Anderson LM, Harris J, Noyes J. Assessing the complexity of interventions within systematic reviews: development, content and use of a new tool (iCAT_SR). BMC Medical Research Methodology 2017; 17: 76.
- Shepherd J, Kavanagh J, Picot J, Cooper K, Harden A, Barnett-Page E, Jones J, Clegg A, Hartwell D, Frampton GK, Price A. The effectiveness and cost-effectiveness of behavioural interventions for the prevention of sexually transmitted infections in young people aged 13-19: a systematic review and economic evaluation. Health Technology Assessment 2010; 14: 1-206, iii-iv.
- National Institute for Health Care Excellence. NICE Process and Methods Guides. Methods for the Development of NICE Public Health Guidance. London: National Institute for Health and Care Excellence (NICE); 2012.
- Thomas J, Harden A. Methods for the thematic synthesis of qualitative research in systematic reviews. BMC Medical Research Methodology 2008; 8: 45.
- Lewin S, Booth A, Glenton C, Munthe-Kaas H, Rashidian A, Wainwright M, Bohren MA, Tuncalp O, Colvin CJ, Garside R, Carlsen B, Langlois EV, Noyes J. Applying GRADE-CERQual to qualitative evidence synthesis findings: introduction to the series. Implementation Science 2018; 13: 2.
- Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, Altman DG, Barbour V, Macdonald H, Johnston M, Lamb SE, Dixon-Woods M, McCulloch P, Wyatt JC, Chan AW, Michie S. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. BMJ 2014; 348: g1687.
- Greenhalgh T, Kristjansson E, Robinson V. Realist review to understand the efficacy of school feeding programmes. BMJ 2007; 335: 858.
- France EF, Ring N, Noyes J, Maxwell M, Jepson R, Duncan E, Turley R, Jones D, Uny I. Protocol-developing meta-ethnography reporting guidelines (eMERGe). BMC Medical Research Methodology 2015; 15: 103.
- Tong A, Flemming K, McInnes E, Oliver S, Craig J. Enhancing transparency in reporting the synthesis of qualitative research: ENTREQ. BMC Medical Research Methodology 2012; 12: 181.

제23장 특수 무작위배정 임상시험 포함

핵심 포인트

- 군집 무작위배정 임상시험 및 교차 임상시험과 같은 비표준 설계는 해당 설계에 적합한 방법을 사용하여 분석해야 한다.
- 문헌고찰에 포함된 연구의 저자가 설계 때문에 발생하는 중재결과 자료 간의 상관관계를 설명하지 못할 경우, 문헌고찰 저자가 대략적인 방법을 적용할 수 있다.
- 비뚤림위험 평가 도구의 확장판으로 군집 무작위배정 임상시험에 사용할 수 있다. 개별 연구대상자를 군집 내에서 식별하고 모집하는 방법에서 발생하는 비뚤림 가능성에 각별히 주의해야 한다.
- 교차 임상시험에 대해 비뚤림위험 평가 확장판을 이용할 수 있다. 잔여효과(carry-over effect)로 인한 비뚤림위험에 주의를 기울여야 한다. 또한 '기간 효과(period effect)'의 가능성에도 각별히 주의해야 한다.

23.1 군집 무작위배정 임상시험

23.1.1 소개

군집 무작위배정 시험에서는 개인보다는 개별 군에 다른 중재가 무작위로 배정된다. 이 때 '배정의 단위'를 군집(cluster), 혹은 군이라고 부른다. 그 군은 학교, 마을, 진료소, 가족일 수 있다. 군집 무작위배정 임상시험이 시행되는 몇 가지 이유가 있다. 우선 중재의 집단에 대한 효과, 예를 들어 백신의 집단 면역성을 평가해야 할 수 있다. 다른 이유로 연구대상자가 동일한 환경에서 관리될 때 중재에 대해 '오염'되는 것이 있다. 클리닉에서 임상의의 훈련을 평가하는 임상시험이 그 예가 될 것이다. 단순히 편의를 위해 군집 무작위배정 설계를 하는 경우도 있다.

군집 설계를 하게 되면 한 군집 내의 연구대상자들이 유사한 방식으로 응답하기 때문에 그들의 자료는 더 이상 독립적이라고 가정할 수 없다. 군집 무작위배정 임상시험을 분석할 때는 이 문제를 고려하는 것이 중요하다. 안타깝게도, 과거에 많은 연구들이 배정 단위가 개별 연구대상자인 것처럼 잘못 분석되었다. 분석 단위는 배정(allocation) 단위와 다르며 이를 혼동하여 생기는 오류를 '분석 단위 오류'라고 한다. 군집이 무시되고 개인이 무작위배정 된 것처럼

분석하면 신뢰 구간은 인위적으로 좁아지고, P 값은 인위적으로 작아진다. 이것은 그 중재가 효과가 있다는 잘못된 결론을 초래할 수 있다. 메타 분석의 맥락에서, 군집이 무시된 연구는 더 많은 가중치를 받게 될 것이다.

어떤 경우 개인은 자신의 여러 부분(예: 양쪽 눈이나 여러 개의 치아)에 적용되는 중재에 배정되거나, 한 연구대상자에 대해 반복적인 관찰이 이루어진다. 이러한 경우는 한 개인이 진료소에서 군집될 수 있는 것과 동일한 방식으로 개인 내에 군집된다. 자료가 연구대상자 내에서 군집되는 것을 고려하지 않고 개별 단위(예: 각 치아 또는 각 관측치)에 의해 분석이 이루어진 경우 분석 단위 오류가 발생할 수 있다.

23.1.2 군집 무작위배정 임상시험의 비뚤림위험 평가

군집 무작위배정 임상시험과 개별 무작위배정 임상시험의 주요 차이점은 관심 있는 개인(군집 내)이 한 가지 중재에 직접 배정되지 않을 수 있다는 것이다. 특히, 때때로 개인을 군집에 배정하고 나서 연구에 모집 또는 다른 방법으로 분석에 포함하도록 선택)하는 경우도 있어, 개인을 분석에 모집하는지 또는 선택하는지 여부에 영향을 미치는 배정에 대한 지식이 있을 가능성이 생길 수 있다. 이러한 일이 발생했을 때 발생하는 비뚤림을 연구대상자 식별/모집 비뚤림(identification/recruitment bias)이라고 한다.

군집 무작위배정 임상시험과 개별 무작위배정 임상시험의 두 번째 주요 차이점은 군집 무작위배정 임상시험에서 '연구대상자'를 식별하는 것이 항상 간단하지 않다는 것이다. 그 이유는 다음과 같다.

1. 일부 임상시험에서 연구대상자를 정식으로 모집하지 않을 수 있다.
2. 어떤 임상시험에서는 서로 다른 중재결과를 측정하는 두 개 이상의 연구대상자 군이 있을 수 있다(예: 임상의와 환자에게 측정된 결과).
3. 일부 임상시험에서는 서로 다른 개인에 대해 두 개 이상의 시점에서 자료를 수집한다(예: 설문조사를 통해 지역사회 신체활동 측정, 기저 상태 및 중재 후 평가 연구대상자가 다름).

군집 무작위배정 임상시험에 대해 RoB 2 도구를 사용하여 비뚤림위험 평가를 할 때 연구대상자는 조사자가 관심 중재결과를 측정하고자 하는 자로 정의된다.

RoB 2 도구 중 군집 무작위배정 임상시험용 확장판이 있다(표 23-1). 이 도구는 군집 무작위배정 임상시험에 초점을 맞추어 비뚤림위험을 평가한다. 대부분의 군집 무작위배정 임상시험은 본질적으로 실용적이며 의료에 대한 높은 수준의 의사결정을 지원하는 것을 목표로 하기 때문에, 도구는 현재 중재배정 효과(중재준수 효과 제외)만을 고려한다.

표 23-1. 군집 무작위배정 임상시험 ROB 도구 평가에서 다루어지는 이슈

1. 무작위배정 과정에서 발생하는 비뚤림

- 군집 무작위배정 프로세스는 다양하다. 군집에 대해 순차적으로 혹은 한꺼번에 무작위배정 할 수 있다. 최소화가 흔히 사용되며 무작위배정과 동일하게 다루어진다. 군집 무작위배정은 종종 방법론 전문가에 의해 1회 수행되는 경우가 많다.
- 군집 수는 상대적으로 적을 수 있으므로 개별 무작위배정 임상시험보다 확률 불균형이 더 흔하다. 그러한 우연한 불균형을 비뚤림위험의 증거로 해석되어서는 안 된다. 연구대상자 식별과 모집 과정에서 발생되는 비뚤림은 군집 무작위배정 임상시험에만 한정된다.
- 무작위배정 시점과 관련하여 개별 연구대상자를 파악하고 모집한 시점을 고려하는 것이 중요하다
- 군집 무작위배정 후 연구대상자 식별/모집이 발생한 경우, 모집은 중재에 대한 지식에 영향을 받아 비뚤림이 생길 수 있다.
- 연구대상자의 특성(군집이 아닌)의 기저상태 불균형은 식별/모집 비뚤림의 문제를 제시할 수 있다.

2. 의도한 중재로부터 이탈로 인한 비뚤림

- 연구대상자가 자신이 임상시험 중임을 알지 못하는 경우, 임상시험 상황 때문에 발생하는 의도한 중재에서 발생하는 이탈은 없다고 판단한다.
- 연구대상자, 보호자 또는 중재를 제공하는 사람들이 배정된 중재를 알고 있다면, 개별 무작위배정 임상시험과 동일하게 판단한다.

3. 중재결과 자료 결측에 의한 비뚤림

- 군집 또는 군집 내의 개인에 대한 자료가 결측될 수 있다.
- 결측된 자료의 유형 중 하나를 다룰 때 고려해야 할 사항은 개별 무작위배정 임상시험과 동일하며, 두 가지 모두를 다루어야 한다.

4. 중재결과 측정의 비뚤림

- 중재결과 평가자가 연구가 진행 중임을 인지하지 못할 경우, 중재 배정에 의해 평가의 영향을 받지 않아야 한다.
- 중재결과 평가자가 배정된 중재에 대해 알고 있는 경우, 문제는 개별 무작위배정 임상시험과 동일하다.

5. 선택적 연구 결과 보고로 인한 비뚤림

- 개별 무작위배정 임상시험과 같다

23.1.3 군집 무작위배정 임상시험의 분석 방법

군집 무작위배정 임상시험에서 분석 단위 오류를 방지하는 한 가지 방법은 배정과 동일한 수준에서 분석을 수행하는 것이다. 즉, 각 군집의 요약 측정을 사용하여 각 군집이 하나의 개별 군집인 것처럼 자료를 분석할 수 있다. 그러면 분석의 표본 크기는 군집 수이다. 그러나 이 전략은 군집 크기가 다양할 경우 효과 추정의 정밀도를 불필요하게 줄일 수 있다.

또는 자료의 군집화를 설명하는 방법을 사용하지 않는 한 개인 수준의 통계적 분석은 부적절하게 높은 수준의 정밀도를 초래할 수 있다. 군집 무작위 평가에서 추출 가능한 이상적인 정보는 군집 설계를 적절하게 설명하는 분석에서 필요한 효과 측정값(예: 신뢰 구간과의 오즈비)을 직접 추정하는 것이다. 그러한 분석은 다층 모형에 기초하거나 다른 기법 중에서 일반화된 추정 방정식을 사용할 수 있다. 사용된 방법이 적절한지 여부를 판단하기 위해 통계적 조언을 권고한다. 연구 저자가 그러한 분석을 수행하지 않은 경우, 문헌고찰 저자가 결과를 조정하는 데 사용할 수 있는 대략적인 접근방식은 다음의 두 가지 방법이 있다.

23.1.4 메타 분석을 위한 군집 무작위배정 임상시험의 근사 분석(Approximate analyses) : 유효 표본 크기

과거에 많은 군집 무작위배정 임상시험들은 일반적으로 군집이 아닌 개인에 대해 무작위배정가 수행된 것처럼 분석하여 적절한 중재결과를 보고하지 못했다. 다음과 같은 정보를 추출할 수 있는 경우에는 정확한 분석을 수행할 수 있다.

- 각 중재군에 무작위배정 된 군집(또는 군)의 수 및 연구의 총 연구대상자 수 또는 각 군집의 평균(평균) 크기
- 전체 개인 수에 대해 군집 설계를 무시하는 중재결과 자료(예: 이벤트 또는 연속 자료에 대한 평균과 표준 편차);
- 군집 내(또는 내부) 상관 계수(ICC, Incluster Correlation Coefficiency)의 추정치

ICC는 군집 내 및 군집 간의 상대적 변동성을 추정하는 것이다. 동일한 군집 내 개인의 '동일성'을 설명한다. 모든 결과 보고서에서 ICC를 보고할 것을 권고함에도 불구하고, ICC 추정치는 발표된 보고서에 없는 경우가 많다.

문헌고찰 저자에게 도움이 될 수 있는 일반적인 접근방식으로 유사한 연구에서 얻은 외부 추정치를 사용하는 것이 있다. ICC의 예를 제공하거나 특정 유형의 군집 또는 중재결과에 대해 ICC의 알려진 패턴을 기반으로 한 추정치를 사용할 수 있게 해주는 몇 가지 자원이 있다(Ukoumunne et al 1999). ICC는 다른 유형의 상관관계에 비해 작게 보일 수 있다. 즉 0.05보다 낮은 경우도 많다. 그러나 작은 값이라도 특히 군집 크기가 큰 경우 신뢰 구간 폭(따라서

메타 분석의 가중치)에 상당한 영향을 미칠 수 있다. 경험적 연구에 의하면 자연적으로 더 큰 경향이 있는 군집이 더 작은 ICC를 보였다. 예를 들어, 같은 중재결과의 경우, 지역은 도시나 가족에 비해 ICC가 작을 가능성이 높다.

23.1.4.1 군집 무작위배정 임상시험 통합 예제

예를 들어, 10개 학급 295명의 어린이를 중재군으로, 11개 학급 330명의 어린이를 대조군으로 군집 무작위배정 임상시험을 한다고 해보자. 군집을 무시하고 아이들 사이의 성공 횟수는 다음과 같다고 가정하자.

중재군: 63/295 대조군: 84/330.
군집 내 상관 계수 0.02를 신뢰할 수 있는 외부 자료원에서 얻었고, 경험에 비추어 좋은 추정치로 예상한다고 했을 때, 임상시험의 의 평균 군집 크기는(295+330)÷(10+11) = 29.8이다.

임상시험 전체의 설계 효과는 1 + (M - 1) ICC = 1 + (29.8 - 1) × 0.02 = 1.576이다(M은 평균 군집크기). 중재군의 유효 표본 크기는 295 ÷ 1.576 = 187.2이며, 대조군의 경우 330 ÷ 1.576 = 209.4이다.

연구설계 효과를 사건 수(이 경우 성공)에도 적용하면 다음과 같은 수정된 결과가 나온다.

중재군: 40.0/187.2 대조군: 53.3/209.4.

일단 임상시험의 크기가 효과적인 표본 크기로 축소되면 그 자료를 이분법적 결과나 연속적인 결과로서 RevMan과 같은 통계 소프트웨어에 입력될 수 있다. 연구결과를 반올림하는 것을 기억하면서, 예시 평가의 결과는 다음과 같이 입력할 수 있다.

중재군: 40/187 대조군: 53/209.

23.1.5 군집 무작위배정 임상시험을 통합할 때의 이슈

군집 무작위배정 임상시험은 원칙적으로 동일한 메타 분석에서 개별 무작위배정 임상시험과 결합할 수 있다. 다양한 유형의 임상시험에 대해서 효과의 중요한 차이가 있을 가능성을 고려하여야 한다. 군집 무작위배정 임상시험을 수행해야 하는 좋은 이유가 있으며, 이런 점을 평가하여야 한다. 예를 들어, 감염병 치료에서, 지역사회의 모든 개인에게 적용되는 중재는 재감염의 가능성을 줄일 수 있기 때문에 지역사회 내의 특정 개인에 대한 치료보다 더 효과적이다.

만일 체계적 문헌고찰에 군집 무작위배정 임상시험이 있다면 자료를 어떻게 처리했는지 명시해야 하며 특히, 군집내 상관계수(ICC)가 외부 출처에서 차용된 경우, 결론의 견고성을 조사하기 위해 민감도 분석을 수행해야 한다.

23.1.6 스텝 웨지 임상시험(Stepped-wedge trials)

스텝웨지 임상시험에서 무작위배정은 군집별로 이루어진다. 그러나 실험 중재에 미리 정의된 비율을 배정하고 나머지를 대조군 중재에 배정하는 대신 스텝 웨지 설계는 모든 군집을 대조군에 배정 한 후 순차적으로 일부 군집을 무작위배정을 통해 중재군으로 배정한다. 임상시험이 끝날 때 모든 군집은 중재군에 배정되어 있다. 스텝 웨지 임상시험은 보건 서비스중재와 정책 중재를 평가하는 데 점점 더 많이 사용되고 있으며 모든 군집이 한 번은 중재를 받기 때문에 정책 입안자들에게 매력적이다.

스텝 웨지 임상시험의 자료를 분석할 때는 시간 추세의 가능성을 고려해야 한다. 실험 중재 기간을 비교 중재 기간과 단순히 비교하게 되면 시간에 따라 변화하는 변수에 의해 교란이 생길 수 있다. 왜냐하면 임상시험의 후반기에는 더 많은 군집이 실험 중재를 받기 때문이다.

23.1.7 군집화(clustering)를 사용한 개별 무작위배정 임상시험

개별 무작위배정 된 임상시험에서도 군집화와 관련된 문제가 발생할 수 있다. 이는 동일한 의료인(예: 의사, 외과 간호사)이 여러 연구대상자에게 중재를 제공할 때 발생할 수 있다. 이러한 유형의 군집화는 분석과 관련하여 군집 무작위 임상시험과 유사한 문제를 제기하며, 문헌고찰 저자는 군집 무작위배정 임상시험 평가의 경우처럼 설계 효과를 사용하여 중재 효과 추정치의 분산을 늘릴 수 있다.

23.2 교차 임상시험(Corssover trials)

23.2.1 소개

병렬 그룹 임상시험은 하나 이상의 대체 중재 비교를 위해 각 연구대상자를 단일 중재에 배정한다. 대조적으로 교차 임상시험은 각 연구대상자를 일련의 중재에 배정한다. 단순 무작위배정 교차 설계는 'AB/BA' 설계로 연구대상자를 처음에는 중재 A 또는 중재 B로 무작위배정한 다음, 중재 B 또는 중재 A로 각각 '교차'한다.

교차 설계는 병렬 임상시험에 비해 몇 가지 장점이 있다.

1.각 연구대상자가 자신의 대조군이 되어 연구대상자 간의 차이를 상당히 감소시킨다.

2. 결과적으로, 중재 효과의 추정에서 동일한 정밀도를 얻는데 필요한 연구대상자가 적다.
3. 모든 연구대상자가 모두 중재를 받기 때문에, 개별 연구대상자에 최선의 중재나 선호도가 무엇인지 결정할 수 있다.

어떤 임상시험에서는 개인 내에서 무작위적인 중재가 일어나며, 다른 신체 부위(예: 두 눈이나 양 쪽 치아)에 서로 다른 중재가 적용된다. 신체 부위가 무작위로 분류되고 분석은 개인의 복수 부분(예: 각 눈 또는 입의 각 측면)에 대해 이루어지는 경우, 분석은 교차 임상시험의 분석에서처럼 한 개체에서 쌍을 이룬 것처럼 분석해야 한다.

23.2.2 교차 임상시험의 적합성 평가

교차 임상시험은 안정적이고 만성적인 조건(최소한 연구 대상 기간에 걸쳐)에서 치료의 일시적 영향을 평가할 때 적합하다. 예를 들어 천식, 류마티스성 관절염, 간질 등을 완화하기 위한 중재 연구에 교차 임상시험을 적용할 수 있다. 교차 임상시험이 적절하지 않은 다음과 같은 상황이 있다.

1. 일시적 상태로 임상시험 기간 중 회복될 가능성이 있는 경우, 퇴행성 장애, 순환 장애(cyclic disorder)처럼 시간이 지남에 따라 상태가 진행되는 경우
2. 중재에 의해 영구적 또는 장기적 상태 변화가 생기는 경우(예: 백신). 이 경우, 연구대상자는 다른 중재군으로 들어가지 못하거나 잔여 효과(carry over effects)의 가능성이 매우 높아짐.
3. 약물의 반감기가 매우 길어서 잔여 효과가 있을 가능성이 높은 경우.
4. 약물배출(wash-out)에 의해 금단효과나 반동(rebound) 효과가 생기는 경우.

메타 분석에 교차 임상시험이 포함될 수 있기 때문에 먼저 교차 임상시험이 해당 상황과 중재에 적합한 설계인지의 문제를 다루어야 한다. 예를 들어, 어떤 저자는 퇴행적 특성 때문에 알츠하이머병 연구에 교차 임상시험을 하는 것은 부적절하다고 판단했으며, 체계적 문헌고찰에 교차 임상시험의 첫 번째 기간의 자료만 포함시켰다.

두 번째 문제는 심각한 잔여(carry over) 가능성이 있는지 여부인데, 잔여를 입증하기 위한 통계적 기법이 만족스럽지 못하기 때문에 판단에 주로 의존한다. 중재의 특성과 약물배출 기간은 중요한 고려사항이다.

23.2.3 교차 임상시험의 비뚤림위험 평가

교차 임상시험과 관련된 주요 문제는 잔여이다. 잔여란 한 기간에 주어진 중재의 영향이 후속 기간까지 지속되어 두 번째 중재의 영향을 방해하는 상황을 말한다. 이러한 영향은 첫 번째

중재 자체가 지속되기 때문일 수도 있고(예: 제거 반감기가 긴 약), 중재의 영향이 지속되기 때문일 수도 있다. 극단적인 예로는 관심 주요 중재결과를 되돌릴 수 없거나 장기간인 경우인데 사망이나 수정능력 저하 관련 연구에서 임신과 같은 경우이다. 이 경우 교차 연구는 일반적으로 부적절하다고 할 수 있다. 잔여 효과는 치료법들 사이에 관찰된 중재결과의 차이가 그들이 치료를 받은 순서에 따라 나타날 수 있는 것을 의미한다. 따라서 전반적인 치료 효과 추정에 영향을 받는다(대체로 과소평가되어 무효점을 향한다). 많은 교차 임상시험에는 잔여를 줄이기 위한 수단으로 약물배출기(washout period)를 둔다.

교차 임상시험에서 발생할 수 있는 두 번째 문제는 기간 효과(period effects)이다. 기간 효과는 동일한 치료를 하였을 때 제1기의 반응과 제2기의 반응과 비교했을 때 두 기간 사이의 체계적 차이다. 질병 상태가 시간이 지남에 따라 체계적으로 변화하거나 보건의료 전략과 같은 배경 요인에 변화가 있을 때 발생할 수 있다. AB/BA 설계의 경우, 동일한 수의 연구대상자가 두 개의 중재시술 순서에 따라 무작위배정 되도록 보장하거나 통계 모델에 기간 영향을 포함함으로써 기간 효과를 극복할 수 있다.

교차 임상시험의 세 번째 문제는 첫 번째 기간을 기준으로 이루어진 분석만 보고할 수 있다는 것이다. 교차 임상시험의 첫 번째 기간은 사실상 병렬 군 비교이지만, 첫 번째 기간의 자료만 사용하기로 잔여 검증 결과에 기초하여 결정한다면 비뚤림이 발생할 것이다. 그러한 '2단계 분석'은 신뢰가 떨어지지만 여전히 사용된다. 이는 잔여에 대한 검정이 무작위배정 된 군의 기저상태 차이에 의해 영향을 받아 통계적으로 유의한 결과가 나올 수 있기 때문이다. 이 상황에서 첫 번째 기간 자료만 보고하는 것은 특히 문제가 있다. 특히 연구자들이 2단계 분석 전략을 사용하여 명시적으로 보고한 경우, 첫 번째 기간의 자료만 이용한 교차 임상시험은 비뚤림 위험이 있는 것으로 간주해야 한다.

RoB 2 도구 중 교차 무작위배정 임상시험용 확장판이 있다. 이 도구에서 다루어지고 있는 이슈는 표 23-2에 있다.

표 23-2. 교차 무작위배정 임상시험 ROB 도구 평가에서 다루어지는 이슈

1. 무작위배정 과정에서 발생하는 비뚤림

- 무작위배정 방법을 둘러싼 문제는 병렬 그룹 임상시험과 동일하다.
- 각 중재 순서에 따라 연구대상자 비율이 동일하게 무작위배정 된 경우 기간 효과가 없어진다(결측 자료에 차등이 없는 경우).
- 연구대상자의 비율이 다르게 무작위배정 되는 경우, 비뚤림을 방지하기 위해 기간 효과를 분석에 포함시켜야 한다.
- 무작위배정 과정의 문제를 검토하기 위해 기저상태 차이 분석을 한 경우, 이는 첫 번째 기간의 시점 차이에 기초해야 한다.

2. 의도한 중재로부터 이탈로 인한 비뚤림

- 교차 임상시험 시 비뚤림위험을 평가할 때 잔여가 핵심 관심사이다. 잔여 효과가 두 번째 기간에 측정한 결과에 영향을 미치지 않아야 한다. 기간 사이에 장시간 약물배출기를 두면 이것을 피할 수 있지만 반드시 필요한 것은 아니다. 중요한 고려사항은 잔여 효과가 사라지도록 2차 기간 중재결과 측정 전에 충분한 시간이 경과하는지 여부다.
- 다른 모든 문제는 병렬 그룹 임상시험과 동일하다.

3. 중재결과 자료 결측에 의한 비뚤림

- 문제는 병렬 그룹 시험과 동일하다. 잔여 효과가 사라지기 전에 다음 관찰을 수행한 경우, 다음 관찰 결과의 처분이 특히 문제가 될 수 있다. 교차 임상시험의 일부 분석은(AB/BA 설계의 경우) 어느 한 기간에 결측된 자료가 있는 모든 환자를 자동으로 제외한다.

4. 중재결과 측정의 비뚤림

- 다른 모든 문제는 병렬 그룹 시험과 동일하다.

5. 선택적 연구 결과 보고로 인한 비뚤림

- 추가적인 우려 사항은 잔여에 대한 검정에 기초하여 1차 기간 자료를 선별적으로 보고하는 것이다.

23.2.4 교차 임상시험의 첫 번째 기간만 사용

문헌고찰에서 교차 임상시험이 예상되는 경우 한 가지 방법은 첫 번째 기간의 자료만 사용하도록 처음부터 계획하는 것이다. 교차 임상시험의 첫 번째 중재 기간만 포함하면 연구에서 정보의 절반 이상이 삭제된다.

문헌고찰에서(처음부터) 교차 임상시험의 첫 번째 기간만 검토하고자 하는 경우, 병렬 그룹 무작위배정 임상시험 버전의 RoB 2 도구를 사용해야 한다. 그러나 잔여가 확인된 경우에 첫

번째 기간의 자료만 보고된 경우 선택적 보고의 잠재적 영향에 유의해야 한다. 쌍으로 구성된 분석만 보고하는 시험의 결측(즉, 첫 번째 기간 동안 자료를 별도로 보고하지 않음)은 메타 분석 수준에서 비뚤림을 초래할 수 있다.

23.2.5 교차 임상시험의 분석방법

잔여 효과나 기간 효과가 문제가 되지 않는 경우, 2주기 교차 임상시험의 연속 자료에 대한 적절한 분석은 paired t-test이다. 이것은 '실험 중재(E)'에서 '대조 중재(C)'에 대한 측정값을 각 연구대상자에 대해 별도로 평가한다. 이러한 차이 측정의 평균 및 SE는 효과 추정과 통계적 테스트의 구성요소들이다. 효과 추정치는 일반적인 역분산 접근법(예: RevMan의 경우)을 사용한 메타 분석에 포함될 수 있다.

다음과 같은 자료를 사용할 수 있는 경우 paired t-test로 분석할 수 있다.

- 논문 또는 임상시험 기관과의 서신을 통해 얻은 개별 연구대상자 자료
- 실험 중재(E)과 대조군 중재(C) 측정 사이의 연구대상자 수준 차이의 평균 및 표준 편차 자료 (또는 표준 오차)
- 평균 차이 및 다음 중 하나: (i) paired t-test의 t-통계, (ii) paired t-test의 P 값, (iii) paired analysis의 신뢰 구간
- 각 개체의 일치된 측정치가 식별될 수 있는 한 개별 자료 값을 추출할 수 있는 실험 중재(E) 및 대조군 중재(C)에 대한 측정 그래프

23.2.6 군집 교차 임상시험

군집 교차 임상시험에는 군집 무작위배정 임상시험과 교차 임상시험의 측면이 결합된다. 2주기의 2개의 중재에 대한 군집 교차 임상시험에서 군집은 실험 중재 또는 대조군 중재에 무작위배정 된다. 첫 번째 기간이 끝날 때, 실험 중재 군집은 두 번째 기간 동안 대조군 중재에 교차하고, 대조군 중재 군집은 두 번째 기간 동안 실험 중재에 교차한다. 군집은 두 기간에 동일한 개인을 포함하거나 두 기간에 다른 개인을 포함할 수 있다. 설계는 군집 수준에서 중재가 가장 적절하게 구현되거나 평가되는 상황에 교차 설계의 장점을 도입한다.

군집 교차 임상시험의 분석은 군집 내 중재 기간의 쌍과 군집 내 개인의 유사성을 모두 고려해야 한다. 안타깝게도 많은 시험들이 적절한 분석을 수행하지 않았으므로 통계적 조언을 구하는 것이 필요할 수 있다.

23.3 참고문헌과 읽을거리

- Eldridge S, Kerry S. A Practical Guide to Cluster Randomised Trials in Health Services Research. Chichester (UK): John Wiley & Sons; 2012.
- Freeman PR. The performance of the two-stage analysis of two-treatment, two-period cross-over trials. Statistics in Medicine 1989; 8: 1421–1432.
- Hemming K, Haines TP, Chilton PJ, Girling AJ, Lilford RJ. The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting. BMJ 2015; 350: h391.
- Higgins JPT, Eldridge S, Li T (editors). Chapter 23: Including variants on randomized trials. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.0 (updated July 2019). Cochrane, 2019. Available from www.training.cochrane.org/handbook.
- Ukoumunne OC, Gulliford MC, Chinn S, Sterne JA, Burney PG. Methods for evaluating area-wide and organisation-based interventions in health and health care: a systematic review. Health Technology Assessment 1999; 3: 5.

제24장 유해효과

핵심 포인트

- 균형 잡힌 관점에 도달하기 위하여, 모든 체계적 문헌고찰은 중재의 유해효과를 고려해야 한다.
- 이 장에서는 유해효과에 대한 문헌고찰 과정을 다음과 같이 설명한다. 1) 유해효과를 바라보는 접근법 설정, 2) 연구설계 선택, 3) 관심 유해효과 선택, 4) 근거검색, 5) 유해효과에 대한 비뚤림 위험 평가, 6) 근거의 합성과 해석
- 유해효과에 대한 상세한 분석은 해를 입힐 가능성에 대한 근거가 치료법이나 정책 결정에 중요한 영향을 미칠 때 특히 중요하다.
- 유해효과 자료는 종종 연구의 주요 이득 중재결과(primary beneficial outcomes)보다 덜 엄격하게 처리되기 때문에, 문헌고찰 저자는 자료를 합성할 때 빈약한 사례 정의, 부적절한 모니터링 및 불완전한 보고의 가능성을 인지해야 한다.

연구 보고서에서 유해효과 용어가 불충분하게 표준화된 채로 사용되어 문헌고찰 저자에게 어려움을 줄 수 있다. 보편적이고 밀접하게 관련된 용어로는 유해사건(adverse event), 유해효과(adverse effect), 중대한 유해사건(serious adverse event), 중대한 유해효과(serious adverse effects), 유해 약물반응(adverse drug reaction), 부작용(side effect), 합병증 및 위해(harms) 등이 있다. 이 장에서 우리는 약물사용 또는 기타 중재의 도중 또는 후에 발생하는 바람직하지 않거나 유해한 결과에 대해 유해사건(adverse event)이라는 용어를 사용하며, 중재와 사건 사이의 인과관계가 적어도 합리적인 가능성이 있는 이상 경우 유해 효과(또는 위해)라는 용어를 사용한다.

유해효과 문제를 다룰 때 몇 가지 특별한 문제에 대해 논의한다. 첫째, 유해효과가 어떻게 정의되고, 확인되고, 분석되고 보고되는지에 관해 연구들 간 광범위한 다양성이 존재할 수 있다. 유해효과에 대한 정보를 수집하는 데는 크게 두 가지 전략이 있다. 연구자는 일반적으로 자료 수집, 사례 정의 및 판정을 위한 프로토콜 정의 절차에 따라 사전에 정의된 유해효과에 대한 조사를 시행하는 **적극적인 모니터링 또는 감시**를 사용할 수 있다. 예를 들어, 관심 사건이 심근 경색인 경우, 연구 프로토콜에 기록된 대로 의심되는 사건에 대한 검사실 및 심전도 자료를 통해 진단할 수 있다. 이들 중재결과는 또한 독립적인 패널에 의뢰되어 사건 발생을 판정하거나 확증할 수 있다. 이러한 능동적인 모니터링은 대개 중재와 관련 있는 것으로 의심되거나 알려진 잠재적 유해사건인 경우에 수행한다.

일관성 없는 결과 정의와 빈약한 확인은 출판된 자료에만 의존하는 문헌고찰에서 문제가 된

다. 발표된 보고서에서 얻은 정보가 불완전하거나 특이성이 부족할 수 있다. 예를 들어, 출판된 연구의 유해효과는 미발표 임상시험보고서에서 보고된 사건에 비해 발병률이 낮고 덜 심각한 것으로 분류되었다. 임상시험보고서에서 임상시험 연구자에 의해 보고된 총 유해효과 건수 중 3%~33%만이 관련 저널 논문에 보고되었다.

둘째, 연구가 계획되었을 때 유해효과가 알려지지 않을 수 있으므로 자료 수집 과정 및 분석 전략이 제대로 적용되지 않을 수 있다. 전향적인 유해효과 자료 수집이 바람직하긴 하지만, 많은 유해효과는 아직 알려지지 않거나 중재와 관련이 있기 때문에 사전에 구체화할 수 없다. 따라서 **자발적 보고 모니터링**이 가능하다면, 연구기간 동안 모든 유해효과(사전 정의된 것 또는 그렇지 않은 것)을 조사할 수 있다. 유해효과를 인식한 연구대상자와 연구원 모두는 언제든지 보고서를 제출할 수 있다. 이런 방법으로 이전에 중재와 관련되지 않았다고 생각했던 새롭고 예기치 않은 유해효과가 발견할 수 있다.

규제 약물(예: 약물, 생물학적 제제, 백신)의 경우, 자발적으로 보고된 유해효과는 사전에 설정된 사전(dictionaries)에 따라 코딩, 그룹화 되고 분류된다. 유해효과 정보 수집에 사용되는 모니터링 방법에 관계없이, 연구자는 유해효과를 **전체 또는 복합 측정**으로 결합할 수 있으며, 이는 심각한 유해효과의 총 수 또는 유해효과로 인한 탈락 수 또는 기관계(예. 종종 위장관, 심혈관계에서의 유해효과)별 총 수로 보고된다. 그러나 이러한 복합 측정(composite measure)은 사건이 정확히 무엇인지에 대한 정보를 제공하지 않으므로, 오심이나 발진과 같은 별개의 또는 개별적인 이상 반응에 대한 세부 정보를 상세화해야 한다. 이상적으로, 유해효과의 정의와 확인은 문헌고찰에 포함된 연구들 간에 가능한 한 균일해야한다.

셋째, 많은 유해효과가 무작위임상시험에서 관찰되기에는 너무 드물거나 오래 걸린다. 일부 유해효과는 거의 발생하지 않거나 중재 시작 후 오랜 기간 후에 나타날 수 있다. 단기간의 추적 관찰만 있는 작은 무작위배정 임상시험은 흔하고 즉각적인 유해효과(예 : 주사 후 피부반응)는 적절하게 포착 할 수 있지만, 대규모 코호트 연구 또는 환자 - 대조군 연구와 같은 비 - 무작위 연구에서만 드물거나 장기간의 유해효과가 관찰 될 수 있다. 그러므로 관심 있는 유해효과의 유형에 따라, 문헌고찰 저자는 무작위배정 임상시험의 기간을 넘어선 확장된 근거를 고려해야 할 필요가 있다. 따라서 문헌고찰 저자는 유해효과에 대한 빈도, 비율 및 위험 추정치를 해석하거나 비교할 때 중재결과 정의 및 모니터링 방법에 세심한 주의를 기울여야한다.

중재의 유해효과를 평가하기 위한 출발점은 문헌고찰에서 중재의 이득과 유해효과 모두를 평가할 것인지 아니면 유해효과만을 평가할 것인지를 고려하는 것이다.

문헌고찰 저자는 사전에 구체화된 유해효과 세트('확증적 접근')에만 집중할 것인지 아니면 문헌고찰 수행 중에 확인된 유해효과에 대한 자료를 분석 할 것인지('탐색적 접근')를 결정할 필요가 있다. 실제로 일부 문헌고찰 저자는 이 두 가지 접근 방식의 하이브리드를 사용한다. 이득과 유해효과를 조사하기 위해 동일한 근거 자료가 사용될 것인지, 유해효과를 조사하기 위해 추가적인 유형의 근거가 요구되는지 여부에 대한 고려도 필요하다. 마지막으로 유해효과에 대한 구체적인 선택과 정의가 고려될 필요가 있다.

24.1 유해효과에 대한 접근법

24.1.1 확증적 접근법(Confirmatory approach)

확증적 접근법에서 문헌고찰 저자는 문헌고찰 프로토콜에서 관심 중재결과로 하나 이상의 유해효과들을 나열한다. Golder와 동료들은 1994년과 2011년 사이에 발표된 유해효과에 관한 체계적 문헌고찰의 약 80 %가 이러한 접근 방식을 사용하여 특정 사건 또는 사건 범주를 주요 관심사로 선정한 이 접근법을 사용했다는 것을 알게 되었다.

확증적 접근법을 채택할 때, 문헌고찰 저자는 중재와 관련될 것으로 예상되거나 이미 인정된 유해효과를 사전에 구체화하는 것을 목표로 해야 하며 정기적으로 그리고 일관되게 연구에서 측정되는 것으로 가정해야 한다. 관심 유해효과의 선택은 생물학적, 생리학적 또는 심리적 타당성에 근거할 수 있다. 예를 들어, 외과적 중재의 문헌고찰에서 '상처 감염'을 유해효과로 사전에 지정하는 것은 타당하다. 유사하게, 혈소판이나 응고에 영향을 미치는 약물 요법에 대한 체계적 문헌고찰은 사전에 구체화된 출혈을 관심 결과로 선정하는 것은 정당할 수 있다. 일부의 경우, 개연성 있는 효과 기전이 아직 확립되지는 않았지만, 이전에 확립된 관찰이나 연관성을 토대로 측정에 대한 유해효과를 선택하는 것이 합리적일 수도 있다. 확증적 접근의 핵심 제한점은 포함된 연구에서 보고된 예기치 않은 유해효과를 처리 할 수 없다는 점이다.

24.1.2 탐색적 접근법(Exploratory approach)

유해효과 문헌고찰에 대한 탐색적 접근법은 관심 있는 특정 유해효과의 사전 구체화를 포함하지 않는다. 오히려 일반적으로 포함 된 연구에서 발견된 이상 반응 자료의 일부 또는 전부를 추출하는 것이 포함된다. 유해효과에 대한 평가의 약 20%만이 이를 주요 접근법으로 사용하고 있다(Golder et al 2013).

이러한 탐색적 접근은 중재의 예기치 않은 희귀한 유해효과를 식별 할 수 있다. 이것은 확증적 접근법을 사용하는 미리 지정된 유해효과에 대한 향후 문헌고찰에서 어떤 중재결과가 조사될 것인지에 대한 정보를 제공할 수 있다. 또한 탐색적 접근법은 기존 안전 프로파일에 추가할 새로운 신호를 생성하는 데 사용할 수 있는, 관찰된 유해효과 목록과 중재 사이의 가능한 연관성에 대한 자료를 제공 할 수 있다.

탐색적 접근법의 한계는 보고된 특정 유해효과(예 : 임상적 중요성보다는 통계적 중요성에 근거한)이 발견된 특성 때문에 보고되었을 수도 있다는 것이다. 또한 문헌고찰 저자가 긴 유해효과 목록을 만날 때, 자료추출 및 분석에 관한 사후 또는 임의의 분석적 결정이 종종 요구된다. 그러한 자료의 선택 및 합성을 위한 절차는 관심 결과가 완전히 지정되지 않은 경우에도 문헌고찰 프로토콜에서 고려할 필요가 있다.

24.1.3 하이브리드 접근법(Hybrid approach)

하이브리드 접근법은 이전에 인지되지 않은, 예상되는 중재의 유해효과를 포착하기 위한 확증적 접근법과 탐색적 접근법의 요소를 결합한다. 이 접근법에 근거한 문헌고찰은 프로토콜에 적은 수의 관심 유해효과를 나열하지만, 조사된 연구에서 이용 가능한 이상 반응 자료를 포착하기 위한 사후 탐색적 분석을 허용한다. 예가 Box 24-1에 나와 있다.

채택된 접근 방식에 관계없이, 문헌고찰 저자는 문헌고찰이 출판된 자료에만 근거할 때 유해효과의 정의와 확인에 문제가 있을 가능성을 염두에 두어야 한다.

Box 24-1. 특정 중재의 유해효과를 검토하기 위한 세 가지 접근법의 예 : 침술

- 확증적 접근법 : 문헌고찰 저자는 사전에 구체화된 피부 감염의 유해효과 및 바늘 삽입시 통증에 대한 자료를 합성하는 것을 목표로 한다.
- 탐색적 접근 : 문헌고찰 저자는 포함된 연구에서 언급된 모든 유해효과에 대한 자료를 합성하는 것을 목표로 한다.
- 하이브리드 접근법 : 문헌고찰 저자는 포함된 연구에서 발견된 다른 유해효과와 함께, 사전에 구체화된 유해효과로 인한 피부 감염 결과 및 총 탈락자 수에 대한 자료를 합성하는 것을 목표로 한다.

동일 문헌고찰에서 중재의 유익하고 부정적인 효과를 모두 검토 할 때 문헌고찰 저자는 유해효과 자료를 식별하려고 할 때 다른 자격 기준을 적용 할 수 있다. 자격 요건이 다를 수 있는 두 가지 주요 측면은 연구 유형과 연구대상자 유형이다. 다른 목적으로 수행된 연구가 고찰의 유해효과 요소에 적합할 수도 있다.

- 다른 연구 유형: 유해효과를 다루기 위해서는 무작위배정 임상시험에서 연구대상자의 규모, 기간 또는 제한된 자격기준으로 인해 효과가 보이지 않을 수 있으므로 무작위배정 되지 않은 연구를 찾아야 할 수도 있다. 24.2절을 참조한다.
- 다른 유형의 연구대상자: 유해효과 자료는 동일하거나 유사한 중재를 평가하지만 다른 집단이나 질병에서 수행된 무작위배정 임상시험에서 얻을 수 있다.
- 다른 목적: 문헌고찰에 관심 있는 연구대상자에 대한 유해효과 자료를 가진 무작위 임상시험일 수 있지만, 문헌고찰과 관련된 유익한 중재결과를 측정하지는 않았다(예 : 질환자의 약물 농도를 평가하는 약동학 연구).

서로 다른 적격 기준에서 유익하고 부정적인 효과(beneficial and adverse effects)를 다루기 위해, 종종 두 개 이상의 연구 세트(24.4절 참조)에 대해 별도의 검색을 수행해할 수 있고,

비뚤림위험 평가를 다르게 계획해야 할 수도 있다.

24.2 연구설계 유형 선택

체계적 문헌고찰은 무작위배정이 중재군간 알려진 그리고 알려지지 않은 교란을 동일하게 분포시켜주기 때문에 일반적으로 무작위배정 임상시험을 포함한다. 그러나 무작위배정 임상시험에서 연구대상자를 선정하는 기준은 위해의 위험이 높은 연구대상자(예: 허약 노인이나 동반질환을 가진 사람)가 제외될 수 있고, 무작위배정 임상시험에서 추적조사 기간이 장기 유해효과를 포착하기에 충분하지 않을 수 있다. 또한 무작위 교차 임상시험은 일부 유해효과에 대한 조사에 적합하지 않을 수 있다. 특히 한 기간의 중재에 노출된 결과로 다른 기간 동안 유해효과가 발생하는 경우에 그렇다. (질병 또는 약물 /의료기기 레지스트리에서 수집된) 코호트 연구와 환자-대조군 연구와 같은 비무작위배정 연구는 무작위배정 임상시험보다 일부 유형의 유해효과에 대한 자료를 제공 할 가능성이 높다. 그러나 비무작위배정 연구는 비뚤림의 위험이 더 큰 경향이 있다.

자발적인 사례 보고 또는 사례 연구는 드물고 이전에 알려지지 않은 사건을 알리는 데 도움이 될 수 있다. 그러나 대부분의 체계적 문헌고찰의 경우 이러한 자료원은(특히 위험 또는 비율을 추정할 수 있는 분모 자료가 없기 때문에) 범위를 조사하기 위한 목적으로만 사용해야 한다. 이러한 자발적 보고서는 관심 중재결과와 관련되거나 중요한 유해효과를 선택할 필요가 있는 프로토콜 초안 작성에 도움을 줄 수 있다.

24.3 관심 유해효과 선택

문헌고찰 저자는 중증도, 시기 또는 알려진 중재 개입 방식에 근거하여 발생할 수 있는 유해효과 유형에 따라 관심 중재결과를 정의할 수 있다. 관심 있는 유해효과의 사전 세부사항(pre-specification)을 알기 위해 다른 출처의 문헌이 사용될 수 있다. 그러한 것에는 증례보고, 인터넷 포럼 환자 보고서, 주제범위 문헌고찰(scoping review), 규제 승인 제품 정보지(예 : 미국 식품의 약국(Food and Drug Administration)에서 제공) 또는 기타 출처(예 : 영국 국립 처방집, Meyler의 약물 유해효과)에서 임상 관찰 내용이 포함된다.

복합 유해효과는 종종 임상시험에서 보고된다. 일반적인 예로는 “유해효과를 가진 총 연구대상자 수” 또는 “유해효과로 인한 탈락자 수”가 있다. 문헌고찰 저자는 수백 가지의 다양한 사건으로 구성 될 수 있는 복합 유해효과를 해석하는 데 큰 어려움이 있음을 인식해야 한다. 드문 중증 유해효과의 신호가 사소한 유해효과에 가려질 수 있기 때문이다. 또한, 문헌고찰 저자는 탈락에 관한 자료를 다음과 같은 이유로 안전성 또는 허용의 대리 표지자로 해석하는 것을 주의해야 한다.

- 중단 사유의 원인은 복잡하다. 경미하지만 귀찮은 유해효과 때문일 수도 있고, 독성, 효능 부족, 비 의학적 이유 혹은 이들의 조합 때문일 수 있다.
- 철회 및 탈락자 수를 낮게 유지하기 위해서, 임상연구 조건 하에서 환자 및 연구자에 대한 압력을 주어서 더 많은 인구집단이 유해효과를 경험한다는 사실이 반영하지 않을 수 있다.
- 눈가림의 해체는 종종 철회 결정보다 선행한다. 이는 환자의 철회에 대한 중재 효과를 과대평가할 수 있게 한다. 예를 들어, 위약군 환자의 증상은 중단으로 이어질 가능성이 적지만, 반대로 유해효과를 암시하는 증상을 호소한 중재군 환자는 더 쉽게 철회 될 수 있다.

24.4 근거 검색

검색 과정을 고려할 때 문헌고찰 저자는 단일 검색을 수행하여 이득과 위해를 모두 평가할 수 있도록 검색하기로 결정할 수 있다. 그렇다면 검색 전략은 확증적, 탐색적 또는 하이브리드와 같은 선택된 접근법 및 이득과 유해에 대한 적격기준의 차이를 고려하여 설계되어야 한다. 특정 연구설계 또는 연구대상자 유형에 제한되지 않고 충분히 광범위한 경우(예 : 특정 약물 이름 또는 중재를 포함하는 모든 연구를 캡처하는 경우) 단일 검색이 합리적 일 수 있다.

일반적으로, 우리는 특히 유해효과의 영향을 평가하는 연구설계가 효능을 보고하는 연구설계와 다른 경우 유해효과 자료에 대한 별도의 맞춤형 검색을 고려할 것을 권장한다. 효능 또는 효과 연구에 중점을 둔 단일 검색만으로 모든 유해효과에 대한 근거를 포괄적인 방식으로 식별하기에 충분하지는 않다.

전자 생의학 데이터베이스의 다양한 내용 및 색인 기술 때문에, 유해효과에 관한 문헌고찰 검색을 단일 검색원으로 제한하거나, 기본 임상시험연구 데이터베이스의 제한된 조합으로 제한하지 않는 것이 중요하다. MEDLINE에서만 검색하는 것은 권장되지 않는다. 약물 중재의 유해효과를 확인하려면 종종 비약물 중재 또는 의료기기에 관한 문헌고찰에 필요한 것과 다른 검색 방법이 필요하다.

문헌고찰 저자는 유해효과에 대한 미공개 자료원을 검색해야 한다. 미출판 자료원은 동료검토되는 저널 이외의 자료원으로 간주된다. 여기에는 임상 연구 보고서, 임상시험 등록 및 규제 기관 웹 사이트가 포함된다. 'OR' 불리언 함수를 사용하여 특이적인 자연어 검색들과 결합된 일반 색인 용어를 사용하는 것이 좋다. 특이적인 검색 기술과 일반적인 검색 기술에는 모두 장점과 한계가 있지만, 결합하면 장점이 증가하고 제한점이 줄어든다. 색인 용어와 자연어 검색을(가능한 경우) 결합하여, 검색 민감도를 높이고 관련 자료가 누락될 가능성을 줄이는 것이 좋다.

24.5 유해효과에 대한 비뚤림위험 평가

유해효과는 일차 중재결과로 정해지는 경우는 거의 없고, 종종 사전에 구체화되지도 않으므로 유해효과 자료를 확보하는데 사용된 방법에 대한 명확성이 부족할 수 있다. 따라서 유해효

과를 측정, 기록 및 보고하는 방식으로 인해 유해효과는 비뚤림에 취약할 수 있다. 중재결과 측정법은 유해효과를 발견하는데 적절해야 하며, 중재군 전체에 유사한 방법을 사용하여 측정, 확인해야 한다. 중재결과 자료 결측에 의한 비뚤림위험을 평가할 때는 다른 추적기간 및 정보 중도절단(informative censoring)의 가능성을 고려하는 것이 중요하다.

문헌고찰 저자는 현재 권장되는 비뚤림위험 평가도구인 무작위배정연구를 위한 RoB 2 도구 및 비무작위 중재연구를 위한 ROBINS-I 도구를 사용해야 한다. 유해효과 자료에 대한 비뚤림위험을 평가할 때 특히 어려운 점은 선택적 보고 비뚤림이다. 자발적으로 보고된 불리한 중재결과에 근거한 연구결과는 주목할만한 결과에 근거하여 선택되었다는 우려를 가져올 수 있다. 마찬가지로, 특정 연구결과를 숨기거나 강조하기 위해 비정상적인 복합 중재결과로 보고될 수 있다. 유해효과를 누락하거나 부분적으로 보고된 유해효과 자료는 유해효과를 평가하는 체계적 문헌고찰에서 일반적이다. 보고 비뚤림때문에 중재가 안전하거나 상대적으로 유해효과가 없다는 잘못된 결론을 도출할 수 있다.

유해효과 자료에 대한 비뚤림위험을 평가할 때의 주요 이슈는 중재결과 정의 및 유해효과 모니터링 방법이다. 이들은 위해에 대한 빈약한 정의, 확인 보고로 인해 무효점을 향하는 비뚤림이 발생한 중대한 우려가 있는 경우 특별한 주의를 기울여야 한다. 이는 임상시험에 관심있는 중재결과로 미리 구체화되지 않았으며, 모니터링 및 보고 내용이 잠재적으로 부적절할 수 있는 새롭거나 예상치 못한 유해효과에 특히 중요하다. McHarm 도구(Chou et al 2010) 및 Agency of Healthcare Research and Quality(AHRQ) 평가 도구(Chou et al 2007, Viswanathan and Berkman 2012)와 같은 추가 도구들은 이러한 문제들에 대한 추가 내용을 제공한다.

24.6 근거의 합성과 해석

문헌고찰 저자는 유해효과가 연구들에서 사전에 구체화된 중재결과로서 정의되고 모니터링되며, 보고될 때, 유해효과 자료에 대한 해석에 대해 더 큰 확신을 가질 수 있다. 반대로, 유해효과를 예상치 못했거나 자발적인 보고를 통해 사후에 확인된 경우, 문헌고찰 저자는 모니터링 및 보고가 해당 사건을 정확하게 포착할 수 있을 만큼 충분히 강력하다는 근거가 없는 한, 인지된 안전성이나 위해 없음에 대해 보다 신중하게 해석해야 한다(Loke and Mattishent, 2015).

여러 연구들에서 유해효과 자료를 비교 또는 합성하기 전에 포함된 다양한 연구들에서 위해결과에 대한 정의 및 확인 방법이 일관성이 있고 유사한지를 평가하는 것이 중요하다. 유해효과에 대한 효과 추정에서 잠재적 이질성의 중요한 원인은 중재결과 정의 및 측정의 변이이다. 문헌고찰 저자는 연구 저자에게 좀 더 일관성 있게 재분석할 수 있는, 추가 자료 또는 분리 자료 제공을 요청하여 모호성을 해결해야 한다.

복합 측정에서 유해효과를 함께 그룹화하면(예 : 총 유해효과 수), 광범위한 함의를 줄 수 있

는데, 개별 유해효과에 가려진 중재 간에 진정한 차이를 만들어낼 수 있다. 문헌고찰 저자는 연구 사이의 유해효과의 코딩 및 분류의 차이로 인해, 일차 연구에서 보고되지 않은 복합 범주를 구성함으로써 분석에 이용 가능한 사건의 수를 증가시키는 것을 피해야 한다.

반대로, 문헌고찰 저자는 유해효과의 코딩이 자료를 불필요하게 분할하는 상황(예 : 다리 통증, 팔 통증)에 주의해야 하는데, 이는 좀더 전반적인 효과(예 : 통증의 영향을 받는 모든 환자)의 신호를 희석할 수 있다. 문헌고찰 저자는 '결과 요약' 표에 하나 이상의 유해효과 결과를 포함해야 한다. 문헌고찰이 유해효과에 대한 상세한 평가에 중점을 두지 않았다면, 문헌고찰 저자는 중재가 안전한 것으로 보이기(또는 암시)보다는 해가 평가되지 않았다는 명시적 진술을 해야 한다.

연구 보고서에 명확하게 나와 있지 않으면 자료를 '0'으로 추출하지 않는 것이 좋다. 심부전이 명시적으로 '제로'로 보고된 경우에도 문헌고찰 저자는 부정적인 결과에 대한 모니터링의 엄격성(예: 심부전에 대한 특정 능동적 감시 vs. 과소 보고되기 쉬운 자발적 보고에 의존했는지에 대한 세부사항)에 대해 포함된 연구의 방법 부분을 신중하게 확인하는 것이 좋다. 모호성은 종종 위해 없음을 추출하고 해석하는 과정에서 발생하므로, 문헌고찰 저자는 '사건발생 없음(zero events)' 결정에 어떻게 도달했는지 기록해야 한다.

24.7 참고문헌과 읽을거리

- Guy Peryer, Su Golder, Daniela R Junqueira, Sunita Vohra, Yoon Kong Loke; on behalf of the Cochrane Adverse Effects Methods Group. Chapter 19: Adverse effects : Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.0. Cochrane, 2019. Available from https://training.cochrane.org/handbook/current/chapter-19
- McMaster tool for assessing quality of harms assessment and reporting in study reports (McHarm) - Santaguida PL, Raina P. The development of the McHarm quality assessment scale for adverse events: Delphi consensus on important criteria for evaluating harms. http://hiru.mcmaster.ca/epc/mcharm.pdf. Accessed May 14, 2008. Available from https://www.ncbi.nlm.nih.gov/books/NBK47098/table/cerguideharms.t3/
- Golder S, Loke YK, Zorzela L. Some improvements are apparent in identifying adverse effects in systematic reviews from 1994 to 2011. Journal of Clinical Epidemiology 2013; 66: 253-260.
- Viswanathan M, Berkman ND. Development of the RTI item bank on risk of bias and precision of observational studies. Journal of Clinical Epidemiology 2012; 65: 163-178.
- Loke YK, Mattishent K. If nothing happens, is everything all right? Distinguishing genuine reassurance from a false sense of security. CMAJ: Canadian Medical Association Journal 2015; 187: 15-16.

제25장 근거 축적을 위한 전향적 접근

핵심 포인트

- 체계적 문헌고찰은 현재의 지식 상태를 반영해야 하지만, 현재성(currency)를 유지하는 것은 자원의 제한 때문에 어려울 수 있다. 발표된 문헌고찰이 최신이 아닌지 여부를 아는 것이 쉽지는 않지만 업데이트가 필요할 수 있는 시점을 식별하는 데 도움이 되는 도구를 활용할 수 있다.
- 살아있는 체계적 문헌고찰이란 지속적으로 업데이트되는 체계적 문헌고찰로, 새로운 근거가 입수되는 즉시 통합하는 것이다. 특히 연구가 자주 출판되는 빠르게 발전하는 분야에서 유용하다. 새로운 근거의 신속한 식별과 합성을 촉진하기 위해 자료 저장과 재사용을 위한 새로운 기술과 더 나은 프로세스가 개발되고 있다.
- 전향적 메타 분석은 연구결과가 알려지기 전에 메타분석에 적격하다고 확인되거나 적격할 것으로 예상되는 연구에 대한 메타 분석이다. 그들은 보통 포함될 연구의 저자를 포함한 협력 그룹에 의해 수행되며, 그들은 보통 개별 연구대상자 자료를 수집하고 분석한다.

25.1 근거 감시: 근거 축적에 대한 능동적 감시

25.1.1 체계적 문헌고찰의 현재성(currency) 유지

새로운 근거를 통합하여 체계적 문헌고찰의 현재성을 유지하는 것은 많은 경우에 중요하다. 예를 들어, 한 연구에 의하면 문헌고찰의 결론은 5년 이상 유효할 수 있지만, 23%의 연구결과는 2년 이내에 뒤떨어진 정보가 될 수 있으며, 7%는 출판 당시에 뒤떨어진 정보가 될 수 있다(Shojania et al 2007). 빠르게 발전하는 분야의 체계적 문헌고찰은 특히 시대에 뒤떨어질 위험이 있으며, 체계적 문헌고찰이 업데이트 되어야 할 때를 식별하기 위한 다양한 방법이 개발되고 있다.

최근 체계적 문헌고찰의 현재성을 유지하기 위해서는 적어도 2년마다 갱신하는 것이 권고된다. 그러나 그렇게 하는 것은 다음과 같은 이유로 잘 지켜지지 않는다 : 자원의 부족; 새로운 문헌고찰을 시작하는 것과 과거 문헌고찰을 갱신하는 것의 균형을 유지해야 하는 필요성; 일부 의료 분야에서 급격히 증가하는 연구량과 다른 분야에서의 새로운 증거의 부족; 체계적 문헌고찰의 현재성이 떨어져, 해로운 조언을 해 줄 수 있는 위험이 있는지 여부를 파악하기 어려움.

25.1.2 업데이트 신호

업데이트의 우선순위를 정하고 여기에 맞추어 업데이트를 수행하는 전략이 필요하다. 기존에 보고된 연구결과가 변경될 가능성이 있는지 여부에 따라 업데이트를 하며, 업데이트의 잠재적 영향력을 평가할 필요가 있다.

새로운 근거의 가용성에 대한 정보는 다양한 출처에서 나올 수 있으며 다음을 포함한 다양한 접근법을 사용할 수 있다.

- 최초 문헌고찰의 전제 검색 전략 재실행
- 간략한 검색 전략 사용
- 문헌 알림 서비스 이용(using literature notification services)
- 기존 문헌고찰에 포함된 연구보고서를 기반으로 하는 기계 학습 알고리즘 개발
- 임상시험 레지스트리(clinical trials registry) 연구 검색
- 관련 체계적 문헌고찰에 포함된 일차연구 검토
- 기타 공식 감시 방법

문헌 통지 서비스(alerts)를 사용하여 서지 데이터베이스 검색을 간소화할 수 있으며, 이에 따라 검색은 정기적으로 자동으로 실행되며 관련성이 있는 새로운 연구가 제공(push)된다. 또는 응용 프로그램의 프로그래밍 인터페이스를 통해 자동 검색을 실행할 수도 있다. 일부 데이터베이스만 통지 서비스를 제공하기 때문에 제공하지 않는 데이터베이스 중 일부 데이터베이스에 대해서만 개방형 API(application programming interface))를 통해 문헌고찰 저자가 자체 자동 검색을 설정할 수 있다. 이 접근방식은 문헌고찰과 관련이 있을 가능성이 높은 연구가 'push' 모델 내에서 작동할 수 있도록 연구에 대해 체계적으로 색인이 이루어질 때 유용하다. 이러한 접근의 주요 문제점은 임상시험 등록부, 웹사이트, 기타 미출판자료 등은 수기검색을 필요로 하므로, 모든 새로운 근거를 'push' 서비스에만 의존하는 것은 부적절하다는 것이다.

새로운 근거가 체계적 문헌고찰의 결과에 미칠 수 있는 정도를 평가하기 위한 통계적 방법이 제안되었다. 표본 크기 계산를 이용하여 특정 표본 크기의 추가 연구가 업데이트된 메타 분석 결과에 어떤 영향을 미칠 수 있는지에 대한 정보를 얻을 수 있다. 이러한 방법으로 분석해 보면 새로운 근거는 변량 효과 메타 분석에 거의 영향을 미치지 않는다. 단 미래 연구가 기존 연구와 유사할 것이라는 가정이 필요하다. 따라서 체계적 문헌고찰의 업데이트 여부를 결정하는 데 있어 이런 방법을 실제로 사용하는 것은 제한적일 수 있다.

25.1.3 '살아있는' 체계적 문헌고찰

'살아있는' 체계적 문헌고찰(LSR, Living systematic review)은 새로운 근거가 나오면 이를 지속적으로 업데이트하는 것을 말한다. LSR을 위해서는 규칙적으로 그리고 빈번하게 업데이트를 해야 하는데 이 방법론은 문헌고찰 우선순위가 매우 높지만, 근거의 확실성이 낮고, 그리

고 새로운 연구 근거가 나올 가능성이 높을 경우에 제안될 수 있다.

새로운 연구 근거에 대해 잦은 검색(매월)을 통해 근거에 대한 감시가 이루어지며, 새로운 정보가 적시에 문헌고찰에 통합되도록 한다. LSR 계획을 수립할 때 중요한 이슈는 검색방법과 예측된 문헌고찰 업데이트의 빈도를 계획서에 명백히 밝히는 것이다. 이러한 투명성은 최종 사용자에게 유용하며, 새로운 버전의 예상 날짜를 기준으로 중요한 의사결정일을 정할 수 있고, 문헌고찰 업데이트 필요성을 줄일 수도 있다. LRS의 유지보수를 통해 의사결정자는 LSR의 근거 업데이트에 따라 자신의 의사결정도 최근 근거에 기반하여 할 수 있는 기회를 얻게 된다.

25.1.4 근거 감시의 기술(techologies)

더 자주 문헌고찰에 대한 업데이트를 시행하면 현재성 측면에서 이익을 얻을 수 있지만, 기존 접근법보다 훨씬 더 많은 자원을 소비하지 않으려면 검색을 간소화해야 한다. 정보와 컴퓨터 과학 의 발전에 따라 검색의 자동화를 통해 수작업의 감소 가능성이 생겼다.

새로운 시스템(Epistemonikos와 같은 데이터베이스)은 개인들이 검색해야 하는 데이터베이스 수를 줄이고, 문헌고찰팀 작업의 중복을 감소시킨다. 또한 오픈 액세스 출판물에 대한 관심이 증가함에 따라 OpendCitation, CrossRef 및 Microsoft Academic과 같은 오픈 액세스 서지 자료에 대한 대규모 데이터 세트도 생성되었다. 이러한 데이터 세트가 특정 분야의 거의 모든 관련 자료를 포함할 수 있도록 커지면, 수많은 자료원을 검색해야 할 필요성이 줄어들 수 있다.

또한 정기적인 검색을 수행하려면 검색된 기록에 대한 정기적인 적격성 평가가 필요하다. 일단 문헌고찰이 완성이 되고, 초기 검색 후 선별 작업이 이루어졌다면, 업데이트는 문헌고찰 저자의 이전 결정에 기초한 적격성을 '학습'하는 자동화 도구를 사용하여 수동 선별 작업노력을 줄일 수 있다. "코크란의 RCT 분류기"와 같이 보다 다수의 자료를 기반으로 하는 자동화 도구도 이용할 수 있다. 이것은 일련의 자료 세트에서 무작위배정 연구라고 할 수 없는 것들을 필터링하는 목적으로 사용된다. 코크란 연합은 RCT로 분류된 내용을 크라우드에 저장하는 코크란 크라우드(Cochrane Crowd)를 개발했다.

문헌고찰 절차의 후반 단계에도 새로운 기술의 도움을 받을 수 있다. 여기에는 비뚤림위험 평가, pdf 표에서 구조화된 자료 추출, 보고서에서의 정보 추출(연구대상자수, 중재의 특징), 심지어는 문헌고찰 연구결과 작성 등이 포함된다. 하지만 아직은 연구 식별에 사용되는 기술보다는 덜 발달되어 있다.

이러한 다양한 도구는 표준적인 체계적 문헌고찰 과정상의 특정 지점에서 수기 작업을 줄이는 것을 목표로 한다. 코크란 연합은 더 큰 그림을 그리고 있는데 그것은 적절한 출처에서 전향적인 중앙 집중식 루틴 검색을 통해 서지 사항, 임상시험 레지스트리 기록 등을 사전에 확인한다. 이러한 기록에 대해 기계 학습과 인간의 노력(Cochrane Crowd 포함)으로 기록을 세부적으로 분류하는 '파이프라인'을 통과하게 된다. 연구 유형을 결정하고, 무작위배정 임상시험

일 가능성이 있는 경우, 이 레코드는 코크란 온톨로지(ONTOLOGY) 용어를 사용하여 주제가 검토되고, PICO 요소에 따라 분류된다. 마지막으로, 전체 본문에서 관련 자료를 추출한다. 이러한 시스템의 성패는 정확도에 있다. 따라서 우선 RCT에 초점을 맞춘 것은 적절하다고 할 수 있으며 코크란 크라우드의 경우 전체 기계 학습 시스템은 RCT 식별에 대한 정확도는 99%를 넘는다(Wallace et al 2017).

중앙 집중식 연구 발견(centralized study discovery)을 위한 시스템을 구축하는 것은 규모의 경제를 통해 이익을 창출할 가능성이 크다. 예를 들어, 과거에는 동일한 연구에 대한 고

찰이 여러 문헌고찰에 걸쳐 여러 번 이루어졌다. 왜냐하면 이전에는 문헌고찰에 대한 의사결정(현재의 문헌이 RCT인지 여부와 같은)을 공유할 방법이 없었기 때문이다. 만일 주어진 기록에 대한 의사결정이 한 번만 이루어지도록 하면 수작업의 중복이 상당히 감소될 수 있다. 그런 다음 이러한 결정이 코크란 레지스터에 연구를 포함시키는 데 반영되게 되면, 더 효율적으로 검색할 수 있다. 이 시스템은 하나의 기록이 하나의 문헌고찰에 관련될 경우, 서로 다른 적격성을 가진 문헌고찰에 관련될 가능성이 낮다는 것을 알게 됨으로써 이익을 얻게 된다. 궁극적으로, 중앙 데이터베이스에서 이들의 PICO 분류에 대한 단일 검색을 통해 무작위 임상시험을 식별하는 것이 목표가 될 것이며, 기존 검토에 대한 새로운 연구가 자동으로 식별될 수 있을 것이다.

25.2 전향적으로 계획된 메타분석(Prospectively planned meta-analysis)

25.2.1 전향적 메타분석이란?

적절하게 수행된 체계적 문헌고찰은 잠재적으로 적격인 연구를 식별하기 이전에 다루어야 할 문제를 정의한다. 그러나 본래 체계적 문헌고찰은 후향적이라 할 수 있다. 왜냐하면 포함된 연구는 보통 임상시험이 완료되고 연구결과가 보고된 후에 확인되기 때문이다. 전향적 메타분석(PMA, prospective meta-analysis)은 해당 연구의 결과가 알려지기 전에 메타 분석의 적격 여부를 식별, 평가하여 적격성이 결정된 연구에 대한 체계적 문헌고찰 및 메타 분석을 하는 것이다. 대부분의 PMA 경험은 무작위배정 임상시험에 대해 이루어졌으므로 이 섹션에서도 무작위배정 임상시험에 초점을 맞추어 PMA를 설명한다.

PMA는 다음을 통해 개별 연구대상자 자료 또는 집계 자료에 대한 후향적(retrospective) 메타 분석이 가지는 문제를 극복할 수 있다.

1. 개별 임상시험 연구결과에 대한 사전 지식 없이 가설을 지정함(하위군 분석에 대한 가설 포함)
2. 임상시험 선정 기준을 전향적으로 지정함
3. 개별 임상시험 연구 결과를 알기 전에 분석 방법을 선택함(자료 의존적 의사결정에서 발생하는 해석의 잠재적 어려움을 피할 수 있음)

대부분의 PMA는 임상시험이 진행 중일 때 시작되며, 각 참여 임상시험의 대표를 포함하여 협력 단체에 의해 수행된다. 그들은 각각의 개별 연구대상자 자료(IPD, individual participant data)를 수집하기도 하여, 후향적 IPD 메타 분석과 공통점이 있을 수 있다.

IPD에 대한 PMA가 충분히 일찍 시작되면, 임상시험 설계, 자료 수집 및 임상시험 프로세

스가 표준화될 수 있는 기회를 제공한다. 예를 들어, 연구자들은 특정 중재결과를 측정하기 위해 동일한 도구를 사용하는 것에 동의할 수 있고, 각 연구의 동일한 시점에 중재결과를 측정할 수 있다. 어린이 비만 예방을 위한 중재에 대한 코크란 리뷰에서, 일부 중재결과 측정의 다양성과 신뢰성의 문제 때문에 임상시험 전반에서 자료 결합이 어려웠다. 이 PMA는 표준화가 되지 않아 발생한 몇 가지 문제를 해결할 수 있도록 공유 표준 세트를 제안했다(Steinbeck et al 2006).

IPD에 기초한 PMA는 심혈관 질환, 소아 백혈병, 소아 및 청소년 비만, 신생아학 분야에서 이루어졌다. 감염병학의 경우 PMA를 사용할 기회를 여러 번 놓쳤다.

자원이 한정되어 있는 경우, IPD가 아닌 집계 자료에 기초한 체계적 문헌고찰과 메타 분석을 수행하는 것이 여전히 가능할 수 있다. 이런 방법론은 그러나 임상시험의 후반부에 시작되기 때문에, 임상시험 수행을 표준화할 기회가 적다. 그러나 메타 분석에 포함시키기 위해 자료를 조화시키는 것은 가능하다.

25.2.1.1 전향적 메타 분석과 대규모 다기관 시험의 차이점은 무엇인가?

IPD에 기반한 PMA는 다기관 임상시험과 유사하며, 표본 크기를 늘리고, 치료 세팅 및 인구집단을 다양화하며, 여러 세팅에 걸친 중재 효과의 이질성을 조사할 수 있는 능력이 커지는 등의 장점을 가지고 있다. 그러나 다기관 임상시험은 모든 기관에서 단일 프로토콜이 구현되는 반면, PMA는 개별 임상시험이 수행되는 방식에 있어 연구자가 더 큰 유연성을 가질 수 있다.개별 기관에서는 개별 프로토콜을 따르며 개별 프로토콜은 모든 임상시험에 공통인 PMA 프로토콜 요소를 갖도록 한다.

PMA는 실용적이거나 정책적인 이유로 적절한 크기의 단일 임상시험이 불가능할 때 매력적인 대안이 될 수 있다. 또한 동일한 문제를 다루는 두 개 이상의 임상시험이 다른 임상시험의 존재에 대해 알지 못하고 시작된 경우, 유용할 수 있다. 이러한 유사한 임상시험이 확인되면, 연구자들은 그들의 결과를 메타 분석으로 결합하기 위해 전향적으로 계획할 수 있다.

포함된 시험 설계를 다양화해서 일반화 가능성을 개선할 수 있는 것이 PMA의 잠재적 장점이다. 예를 들어 FICSIT(Frailty and Injures: 중재 기법에 대한 협력 연구)는 취약한 노인 인구에서 운동 기반 중재 임상시험 8건에 대한 사전 계획된 메타 분석이다. 8개의 FICSIT 기관은 기관별 종결점과 서로 다른 연구대상자 선정 기준(모든 연구대상자가 고령이었다는 점은 공통)을 사용하여 개별 기관의 자체 중재를 세밀하게 정의했다(Schechtman and Ory 2001).

25.2.1.2 협력

후향적 IPD 메타분석과 마찬가지로 임상시험 연구자들과 협력하는 것이 PMA의 성공에 필수적이다. 협력은 보통 프로젝트를 관리하는 운영 그룹이나 사무국을 통해 이루어진다. 협력은 어떤 연구결과가 알려지기 전에 이루어져야하기 때문에 PMA 협력 노력의 중요한 초점은 연구의 임상시험 인구집단, 설계 및 자료 수집 방법에 대한 합의에 도달하는 데 있다. 이상적으로, 협력

그룹은 모든 임상시험에 걸쳐 적용될 핵심 공통 프로토콜과 자료 항목(조작적 정의 포함)에 대해 합의한다. 개별 임상시험 단위로 부분적인 프로토콜 개정이나 추가 자료 항목이 포함될 수 있지만, 연구자는 이러한 것들이 핵심적인 공통 프로토콜 요소를 손상시키지 않도록 해야 한다.

협력 그룹은 각 연구 그룹으로부터 명시적인 협력 계약을 얻는 것이 바람직하다. 이를 통해 개별 연구자의 실질적인 기여를 장려하고, PMA 개념에 대한 이해를 증진시켜야 한다.

25.2.1.3 IPD와 연구결과의 기밀유지

자료 익명성 및 보안과 관련된 기밀 유지 문제는 IPD 메타 분석과 유사하다. PMA에 특이적인 문제로는 임상시험 완료 시, 자료 처리 방법, 자료와 안전성 모니터링과 관련된 이슈 관리 방법 등이 있다.

25.2.2 전향적 메타분석의 계획서 작성

모든 PMA는 PROSPERO 또는 이와 유사한 레지스트리에 등록되어야 하며 공개적으로 이용할 수 있는 프로토콜을 가지고 있어야 한다. PMA용 프로토콜 개발은 개념적으로 표준적 체계적인 문헌고찰을 위한 프로세스와 유사하다. 그러나 일부 고려사항은 다음과 같이 PMA에 고유한 것이다.

목적, 적격성, 중재결과

체계적인 문헌고찰이나 메타 분석과 마찬가지로 PMA 프로토콜은 임상시험의 목적과 적격성(임상시험 연구대상자, 중재군 및 대조군 포함)을 명시해야 한다. 또한 PMA의 모든 임상시험에서 어떤 중재결과를 측정할지, 그리고 언제 어떻게 측정해야 하는지 명시해야 한다. 또한 하위군 분석 변수의 세부 정보를 지정해야 한다.

검색방법

후향적 체계적 문헌고찰과 마찬가지로, 정밀도를 극대화하기 위해 모든 적격의 진행 중인 임상시험을 식별하기 위한 체계적 검색을 수행해야 한다. 프로토콜은 진행 중이거나 계획된 연구를 식별하거나 참여시키기 위해 잠재적 협력자의 위치를 확인하고 접근하는 방법을 포함하여 PMA 개발에 공통적인 관심을 가진 연구자를 식별하기 위한 노력을 상세히 기술해야 한다.

임상시험의 세부 사항

포함이 결정된 임상시험의 세부 사항을 기술한다. 기술해야 하는 사항으로는 임상시험 등록번호, 예상 연구대상자 수, 일정 등이다. 프로토콜은 각 연구의 적절한 대표자(의뢰사 또는 책임연구자)로부터 협력하기로 한 계약을 획득했는지 여부를 명시해야 한다. 이 프로토콜에는 PMA 연구결과는 각 연구의 자체 자료 모니터링 위원회 이외의 누구에게도 알려지지 않았다는 문구가 포함되어야 한다. 적격 시험으로 확인되었지만 PMA 연구 문제와 관련된 연구결과

가 이미 알려져서 PMA에 포함되지 않은 경우, 이러한 자료를 어떻게 처리할지를 프로토콜에 명시해야 한다. 예를 들어, 이러한 임상시험 자료를 포함한 민감도 분석을 계획할 수 있다.

자료통합 및 분석

프로토콜에 표준 집계 자료 메타 분석 또는 IPD 메타 분석과 유사한 방식으로 자료 수집 및 분석 계획을 개략적으로 설명해야 한다. 또한 전체 표본 크기 및 검정력 계산, 중간 분석 및 하위군 분석의 세부 정보도 제공해야 한다. 전향적으로 계획된 임상 시험이 있는 경우 순차적 접근방식으로 메타 분석을 진행하는 것이 합리적일 수 있다.

IPD-PMA에서, 프로토콜은 PMA 내의 임상시험 연구자들이 연구대상자 수준 자료를 제공할 수 없을 경우 어떻게 할 것인지 기술해야 한다. 프로토콜은 지속적인 자료 수집 주기(예: 5년)를 통해 정기적으로 PMA 자료를 업데이트할 의도가 있는지 여부를 명시해야 한다. 상세한 통계 분석 계획은 PMA에 포함될 자료의 수집 또는 분석 전에 합의하고 공개해야 한다.

25.2.3 전향적 메타분석의 자료 분석

연구대상자 수준의 자료를 사용하면 중재 효과가 연구대상자 특성에 따라 달라지는지에 대한 통계적 검정을 수행할 수 있으며, 어떤 경우에는 예측 모델링이 가능한 경우도 있다.

25.2.4 전향적 메타분석의 중간 분석과 자료 모니터링

개별 임상 시험 자료의 중간 분석 계획, 특히 중재의 안전성을 모니터링하기 위한 계획이 포함되는 경우가 많다. PMA를 통해 모든 자료를 사용하여 중간 분석을 수행할 수 있다. PMA용 DSMC(Data Safety Monitoring Committee)의 지원을 받아 사용 가능한 자료를 모든 임상시험 기관에서 받아 결합하여 중간 분석을 수행하거나 각 임상시험별로 평가한 다음 모든 참여 임상시험 DSMC 사이에 해당 연구결과를 공유할 수 있다.

결합된 중간 분석을 수행할 수 있는 능력에 대해 몇 가지 윤리적 문제가 제기되었다. 예를 들어, 통합 분석에서 중재에 대한 전체적인 순이득이 입증된 경우 개별 임상시험을 계속 진행하는 것이 적절한가? 하위군 또는 종결점 중재결과를 알 수 없는 경우, 연구자는 전체 순이득에 대한 추가 정보를 얻기 위해 PMA를 계속 진행해야 하는가? 각 시험에 자체 DSMC를 보유하고 있는 경우, 위원회 간의 의사소통을 하면 이 상황에서 유용할 것이다. 예를 들어, 결합된 중간 자료의 효과를 근거로 개별 임상시험을 조기에 종료할지 여부를 결정하는 데 도움이 될 수 있다. 모든 참여 임상시험의 자료를 결합한다면 개별 임상시험이 잘못 조기 중단될 가능성을 줄일 수 있다. 따라서 PMA 내의 개별 시험 DSMC는 하위군 혹은 일부 종결점(예들 들어 순이득)에 대해 PMA의 목표가 달성될 때까지 개별 임상시험을 중단해서는 안 된다는 공통의 합의를 채택하는 것이 도움이 될 것이다.

또 다른 가능한 옵션으로 임상적 및 통계적 판단이 타당할 경우 특정 관심 하위군의 연구대상자의 등록을 제한하는 것을 고려할 수 있다. 어떠한 경우든, 개별 시험 결과를 이용할 수 있게 될 때 순차 메타 분석 방법론(sequential meta-analysis methodology)을 적용하는 것이 적절할 수 있다.

25.2.5 집계 자료 기반 전향적 접근 방식: 적응형(adaptive) 메타 분석 프레임워크

적응형 메타 분석(Framework for Adaptive Meta-analysis, FAME)의 틀은 '전통적' 요소와 전향적 요소를 결합한 것으로 집계자료에 적합하며, 곧 나올 임상시험 결과를 포함할 수 있다. 이 접근방식은 완료되었지만 보고되지 않은 임상시험을 포함하여 모든 적격 시험을 고려하는 것을 목표로 하며, 신뢰할 수 있는 집계 자료 메타 분석을 가장 빠르게 얻을 수 있게 해준다. FAME의 주요 단계는 다음과 같다.

1. 대부분의 시험이 진행 중이거나 아직 보고되지 않은 상태에서 체계적 문헌고찰 프로세스를 시작한다. 이를 통해 적격 연구의 결과를 모르는 상태에서 목적, 적격 기준, 중재결과 및 분석을 계획할 수 있으며, 임상시험 연구 결과 출현을 예상하여 문헌고찰 및 메타 분석을 적절하게 조정할 수 있다.
2. 출판, 미출판 문헌 및 진행 중인 연구에 대한 포괄적 검색을 시행한다. 이렇게 하면 메타 분석 계획을 모든 잠재적 임상시험에 기초하여 수립할 수 있고, 현재와 미래 근거의 맥락에서 연구 결과를 만들 수 있다. 초록집, 임상시험 등록처, 연구자 네트워크는 중요한 정보 출처이다. 어떠한 체계적인 문헌고찰이라도, 미출판 자료와 진행중인 연구를 검토해야 하지만, 실제로는 이러한 절차가 표준적이지 않은 경우가 많다(Page et al 2016).
3. 해당 임상시험에 대해 자세히 이해하고 그러한 이해를 유지하기 위해 임상시험 연구자와 연락관계를 유지한다. 가장 중요한 것으로 임상 시험 연구자와의 연락을 통해 임상시험이 어떻게 진행되고 있고 연구결과가 언제 이용가능한지를 파악하는 것이다. 또한 임상시험의 연구 설계, 수행, 분석, 비뚤림위험 평가에 대한 정보도 제공받을 수 있다.
4. 신뢰할만한 메타 분석을 위해 충분한 연구 결과를 사용할 수 있는지 혹은 언제 가능한지를 예측한다. 시간이 지남에 따라 2단계와 3단계 정보를 통해 신뢰할 수 있는 메타 분석이 가능한지와 언제 가능한지를 사전에 평가할 수 있다. 신뢰성의 첫 번째 지표는 메타 분석에 포함될 연구대상자 수와 사건발생 수이며, 이들의 수치가 '최적 정보 크기'를 넘어가는지 살펴볼 수 있다. 두 번째 지표는 예상 연구대상자수와 사건 수가 총 적격자 중 어느 정도 비중을 차지하는지를 보는 것이다(relative information size). 이렇게 하면 보고비뚤림 혹은 자료 이용가능성 비뚤림 위험을 최소화하는 역할을 한다. FAME에 대한 이러한 예측과 결정 사항은 체계적 문헌고찰 프로토콜에 잘 요약되어 있어야 한다.
5. 사용 가능한 자료와 사용할 수 없는 자료를 고려하여 메타 분석을 수행하고 결과를 해석한다. 자료를 해석할 때는 실제 수집된 자료가 얼마나 대표성이 있는지, 그리고 발표되지 않았

거나 진행 중인 시험의 연구결과의 잠재적 영향은 어떠한지를 고려해야 한다. 이는 메타분석의 표준으로 간주되는 메타 분석 결과의 방향과 정밀도에 추가적인 정보를 제공해 준다.

6. 향후 체계적 문헌고찰과 메타 분석에 대한 업데이트의 가치를 평가한다. 메타 분석 연구결과가 결정적인 것으로 간주되지 않으면, 향후 나타날 임상시험 연구 결과를 업데이트하는 것이 가치가 있는지, 그리고 만일 그렇다면 집계 자료로 충분할지 아니면 IPD가 필요한지를 확인할 수 있을 것이다.

25.3 참고문헌과 읽을거리

- Page MJ, Shamseer L, Altman DG, Tetzlaff J, Sampson M, Tricco AC, Catalá-López F, Li L, Reid EK, Sarkis-Onofre R, Moher D. Epidemiology and reporting characteristics of systematic reviews of biomedical research: A cross-sectional study. PLoS Medicine 2016; 13: e1002028.
- Schechtman K, Ory M. The effects of exercise on the quality of life of frail older adults: a preplanned meta-analysis of the FICSIT trials. Annals of Behavioural Medicine 2001; 23: 186-197.
- Shojania KG, Sampson M, Ansari MT, Ji J, Doucette S, Moher D. How quickly do systematic reviews go out of date? A survival analysis. Annals of Internal Medicine 2007; 147: 224-233.
- Steinbeck KS, Baur LA, Morris AM, Ghersi D. A proposed protocol for the development of a register of trials of weight management of childhood overweight and obesity. International Journal of Obesity 2006; 30: 2-5.
- Thomas J, Askie LM, Berlin JA, Elliott JH, Ghersi D, Simmonds M, Takwoingi Y, Tierney JF, Higgins HPT. Chapter 22: Prospective approaches to accumulating evidence. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.0 (updated July 2019). Cochrane, 2019. Available from www.training.cochrane.org/handbook.
- Wallace BC, Noel-Storr A, Marshall IJ, Cohen AM, Smalheiser NR, Thomas J. Identifying reports of randomized controlled trials (RCTs) via a hybrid machine learning and crowdsourcing approach. Journal of the American Medical Informatics Association 2017; 24: 1165-1168.

제26장 문헌고찰 Overview

핵심 포인트

- 문헌고찰 overview(이하 overview)는 중요한 중재결과 전반에 걸쳐 연구결과를 추출 및 분석할 목적으로 유사한 주제에 대해 여러 개의 체계적 문헌고찰을 검색하고 식별하기 위해 명시적이고 체계적인 방법을 사용한다.
- Overview는 중재 문헌고찰과 유사하지만 검색단위, 선택기준과 자료 분석의 대상이 일차 연구가 아닌 체계적 문헌고찰인 점이 다르다.
- Overview는 관심 주제의 체계적 문헌고찰 근거의 현재 상황을 기술하거나, 기존 체계적 문헌고찰에서 수행한 분석과는 다른 방법으로 체계적 문헌고찰 결과 자료를 재분석 할 수 있다.
- Overview는 포함된 체계적 문헌고찰에서 나타난 것과 정확히 일치하는 중재결과 자료를 제시하거나, 체계적 문헌고찰에서 수행된 분석과 다른 방식으로 체계적 문헌고찰 중재결과 자료를 재분석할 수 있다.
- Overview를 수행하기 전에 저자는 overview 형식이 문헌고찰 질문에 적합한지, 그리고 발생할 가능성이 있는 다양한 방법론적 문제를 해결할 준비가 되어 있는지 확인해야 한다.

26.1 서론

이 장은 overview의 정의와 특성, 그리고 overview를 어떻게 구성하고 연구질문을 어떻게 다루며, overview 작성에 필요한 방법론적인 가이드와 절차, 보고 방법 등을 소개한다.

26.2 Overview란 무엇인가?

26.2.1 Overview의 정의

Overview는 주요 중재결과에 대한 각각의 체계적 문헌고찰 연구결과를 추출하고 분석할 목적으로 유사한 주제에 대해 여러 개의 체계적 문헌고찰을 검색하고 확인하기 위해 명시적이고 체계적인 방법을 사용하는 것이라고 정의할 수 있다. 따라서, 검색, 포함기준 및 자료분석의 단위는 체계적 문헌고찰이다. Overview는 일반적으로 다양한 질환의 예방 또는 치료와 관련된 질문(예; 보건의료중재에 대한 질문)에 답하기 위해 수행된다. Overview의 사용자는 보

건의료 의사결정자들이다; 여기에는 의료 제공자, 정책 결정자, 연구원, 자금 제공 기관, 환자 및 보호자, 소비자가 포함된다.

26.2.2 Overview의 구성

Overview는 5가지 요소를 포함해야 한다.

1. 보건의료중재에 대한 특정 연구 질문에 답하도록 정형화된 목표를 가지고 있어야 한다.
2. 체계적 문헌고찰만을 검색하고, 포함하여야 한다.
3. 명시적이고 재현 가능한 방법을 사용하여 overview의 포함기준을 충족하는 여러 개의 체계적 문헌고찰을 확인하고, 이러한 체계적 문헌고찰의 질/비뚤림위험을 평가해야 한다.
4. 포함된 체계적 문헌고찰로부터 다음 자료를 수집, 분석 및 제시해야 한다 : 체계적 문헌고찰과 포함된 일차 연구, 일차 연구의 비뚤림위험, 정량적 중재결과 자료, 사전 정의되고 임상적으로 중요한 중재결과에 대한 근거 수준.
5. overview의 특정 연구 질문과 목적, 목표와 관련하여 확인된 내용들에 대해 다음과 같은 사항을 논의해야 한다 : 주요 중재결과 요약, 근거의 완전성 및 적용 가능성, 근거의 질, overview의 잠재적 비뚤림, 다른 연구 및/또는 체계적 문헌고찰과의 일치 혹은 불일치

26.2.3 Overview에서 다루는 연구질문 유형

Overview는 보건의료 중재와 관련된 5가지 질문의 유형을 다룰 수 있다. 구체적으로, 두 개 이상의 체계적 문헌고찰로부터 다음에 제시된 근거들을 요약할 수 있다 :

- 동일 상황 혹은 인구집단에 대한 다른 중재
- 동일 상황 혹은 인구집단에 대한 같은 중재를 적용한 다른 접근방법
- 다른 상황 혹은 인구집단에 대한 동일 중재
- 동일 상황 혹은 인구집단에 대해 동일 중재이면서, 다른 중재결과 또는 시점이 다른 체계적 문헌고찰을 다루는 경우

표 26-1에서는 이러한 5가지 질문 유형에 관한 추가적인 정보와 예시를 제시하였다.

표 26-1. Overview로 출판하기에 적당한 보건의료 중재에 대한 연구질문의 유형

연구질문의 유형	Overview 예	Comment
동일 상황 또는 인구집단에 대해 서로 다른 중재에 대한 둘 이상의 체계적 문헌고찰에서 근거를 조사	직업 여성을 대상으로 한 통증 관리: (Jones 2012). 보건의료 전문가 행동 및 환자 결과 변화에 대한 재정적 인센티브의 효과(Flodgren 등 2011). 진행성 질환이 있는 성인의 피로 및 체중 감소에 대한 중재(Payne 등, 2012)	이것은 overview에서 다루는 가장 일반적인 질문임
동일 상황 또는 인구집단에 대해 동일한 중재의 적용에 대한 다른 접근방식을 다루는 둘 이상의 체계적 문헌고찰에서 근거를 조사	성인에서 급성 편두통 발병에 서수마트립탄(모든 투약 경로)	이 질문은 종종 overview로 출판하기에 적합함. 이러한 유형의 질문은 투여량, 시기, 빈도, 투여 경로, 기간 또는 투여 과정 수의 차이를 별도의 체계적 문헌고찰에서 다루는 약물 중재에 가장 적합할 수 있음
다른 상황 또는 인구집단에서 동일 중재의 두 개 이상의 체계적 문헌고찰로부터 근거를 조사	소비자의 안전하고 효과적인 의약품 사용을 위한 중재(Ryan 등, 2014). 수술 후 사망 및 주요 이환 예방을 위한 신경 차단술(Guay 등, 2014).	이 질문은 종종 overview로 출판하기에 적합함. 이러한 유형의 질문은 다른 상황이나 인구집단에 걸쳐 동일하거나 유사한 중재의 효과 및/또는 안전성을 검토함
하나 이상의 상황 또는 인구집단에 대한 중재 사용에 대한 둘 이상의 체계적 문헌고찰에서 중재의 부작용에 대한 근거를 조사	천식아동에 대한 포르모테롤 혹은 살메테롤 복용의 안전성(Cates 등, 2012). 성인의 수술 후 급성 통증에 대한 일회성 경구투여 진통제 부작용(Moore 등, 2014).	이 질문은 드물지만 때때로 overview로 출판하기에 적합함. 이러한 유형의 질문은 희귀한 사건의 발생을 확인하고 특성화하는데 도움이 될 수 있음
동일한 상황 또는 다른 결과나 시점을 다른 체계적 문헌고찰에서 다루기 위해 두 개 이상의 동일한 중재에 대한 체계적 문헌고찰에서 근거를 조사	CDSR에는 현재 이러한 유형의 overview가 포함되어 있지 않음	이러한 유형의 질문은 드물지만 때때로 overview로 출판하기에 적합할 수 있음

26.2.4 Overview의 목적

Overview 시행의 목적은 일반적으로 둘 중 하나이다(Pollock 등 (2016), Ballard과 Mont-

gomery (2017)):

목적 1: 관심 있는 특정 문헌고찰 질문에 대해 현재 체계적 문헌고찰 근거를 제시하고 설명한다. 이를 통해 독자들은 명확하게 정의된 관심 주제에 대한 관련 체계적 문헌고찰을 확인할 수 있다. Overview의 목적이 어떤 주제에 대한 체계적 문헌고찰의 현재 근거를 제시하고 설명하는 것이라면, 중재결과 자료는 종종(그러나 항상 그렇지는 않음) 포함된 체계적 문헌고찰 전반을 요약하게 된다.

목적 2: 관심 있는 새로운 혹은 다른 문헌고찰 질문을 다룸. 그 목적은 포함된 체계적 문헌고찰에서 질문하는 것과 달리 2개 이상의 체계적 문헌고찰에서 나온 근거를 이용하여 새로운 문헌고찰 질문이나 다른 문헌고찰 질문에 답하는 것이다. 예를 들어, Overview 질문은 특정 인구집단(예: 성인이 아닌 어린이), 중재(예: 흡입된 치료, 정맥내 치료) 또는 특정 중재결과와 관련될 수 있다. Overview의 목적이 관심 문헌고찰 질문에 대한 답변인 경우, 중재결과 자료는 종종(항상 그렇지는 않지만) 재분석해야 할 필요가 있다(예: 포함된 체계적 문헌고찰에서 중재결과 자료를 추출하고, 체계적 문헌고찰의 분석과 다른 방식으로 자료를 분석하며, Overview에서 재분석된 자료를 제시해야 한다).

26.3 Overview는 언제 시행해야 하는가?

26.3.1 Overview라고 할 수 없는 것은?

Overview에서 하지 말아야할 몇 가지 사항이 있다. Overview는 다음 사항을 목표로 하지 않는다.

- 포함된 체계적문헌고찰의 검색 반복이나 검색 업데이트, 체계적 문헌고찰의 적격성 평가
- 체계적 문헌고찰에서 포함되지 않은 일차 연구에 대한 연구 수준 검색 수행
- Overview 내에 새로운 체계적 문헌고찰 수행
- 체계적 문헌고찰을 시작점으로 하여 일차 연구에서 자료를 추출하고 분석할 목적으로 관련 연구를 찾는 것(이는 overview가 아닌 체계적 문헌고찰 또는 체계적 문헌고찰의 업데이트로 간주됨).
- 체계적 문헌고찰에 대한 사전 정의된 정의를 충족하지 못하는 서술적 고찰, 교과서, 정부 보고서, 임상 진료지침, 기타 다른 요약 보고서를 검색하고 포함하는 것
- 포함된 체계적 문헌고찰의 결론을 추출하여 제시
- 포함된 체계적 문헌고찰에 포함되지 않은 일차 연구의 세부 중재결과 자료 제시

26.3.2 Overview와 중재 문헌고찰 중 선택

체계적 문헌고찰을 검색 대상으로 삼는 경우와 자료 분석을 체계적 문헌고찰 수준에서 할

때만 overview를 실시하는 것이 바람직하다. 체계적 문헌고찰을 검색, 선택하고 분석하면 체계적 문헌고찰보다 범위가 넓은 연구 문제를 다룰 수 있으며, 체계적 문헌고찰이 이미 수행되었기 때문에 시간과 자원이 절약될 수 있다.

일차 연구에 대한 검색을 수행하거나 일차 연구에서 직접 자료를 추출할 필요성을 예상하는 경우 중재 문헌고찰을 수행하는 것이 바람직하다. 또한 저자들이 네트워크 메타 분석을 수행하기를 원한다면 중재 문헌고찰을 실시하는 것이 바람직하다. 간접 비교와 네트워크 메타 분석은 이행성 가정(transitivity assumption)을 철저히 조사하여 그것이 유효한 것으로 판명된 경우에만 수행해야 하는데, 체계적 문헌고찰에서 제공된 정보만을 사용하여 이행성 가정을 적절하게 평가하기는 매우 어렵다.

Overview와 중재 문헌고찰 사이의 주요 차이점은 표 26-2에 요약되어 있다.

표 26-2. Overview와 중재 문헌고찰 사이의 방법 비교

	중재 문헌고찰	문헌고찰 overview
목적	중재의 영향을 조사하는 일차 연구의 근거를 요약하기 위해	중재의 영향을 조사하는 체계적 문헌고찰의 근거를 요약하기 위해
선택 기준	임상적 / 방법론적 선택 및 배제 기준 기술. 관심 있는 연구설계는 일차 연구임	임상적 / 방법론적 선택 및 배제 기준 기술. 관심 있는 연구설계는 체계적 문헌고찰
검색	관련된 일차 연구를 포괄적으로 검색	관련된 체계적 문헌고찰을 포괄적으로 검색
포함	적합성 기준을 충족하는 모든 일차 연구를 포함	적합성 기준을 충족하는 모든 체계적 문헌고찰 포함
방법론적 질/ 비뚤림위험 평가*	포함된 일차 연구의 비뚤림위험 평가	포함된 체계적 문헌고찰의 방법론적 질/ 비뚤림위험 평가. 또한 포함된 체계적 문헌고찰 내의 일차 연구들의 보고 비뚤림위험을 평가
자료 수집	포함된 일차 연구로부터	포함된 체계적 문헌고찰로부터
분석	메타분석, 네트워크 메타 분석 및/혹은 서술적 요약 방법을 사용하여 개별 주요 결과에 대한 포함된 일차 연구들의 결과를 합성	포함된 체계적 문헌고찰 내에 있는 결과 자료를 요약 혹은 재분석
근거의 확실성(예; GRADE)	일차 연구들의 포함된 개별 주요 결과에 대한 분석을 통해 근거 수준 평가	가능하다면, 체계적 문헌고찰에서 제시한 평가내용 보고. 그렇지 않으면 체계적 문헌고찰에 보고된 자료를 사용하여 근거 수준 평가

26.3.3 Overview를 수행하기 전에 고려해야 할 사항

Overview를 수행하는 것이 적절한지 여부를 결정하려면 기존의 체계적 문헌고찰에 대한 약간의 지식이 필요하다. 그런 목적을 위해 체계적 문헌고찰 데이터베이스를 사전 검색해야 한다. 문헌고찰 그룹은 이미 이러한 지식을 보유하고 있을 수 있다.

관심 있는 연구 질문에 대한 overview 형식이 적절한지 여부를 결정하는 8가지 체크 포인트가 있다.

1. 주제는 보건의료 의사결정자 및 문헌고찰 그룹과 관련이 있어야 한다(즉, 보건의료분야에서 의사 결정을 내릴 때 직면하는 실질적인 옵션과 선택사항을 다루어야 함).
2. Overview 포맷을 결정하기 전에 중요한 조직적 요인을 고려해야 한다.
3. 주요 중재 및/또는 관심 질병에 대해 다수의 관련 체계적 문헌고찰이 있어야 한다.
4. 모든 체계적 문헌고찰은 충분히 최신의 것이어야 한다.
5. 모든 체계적 문헌고찰은 인구집단, 중재군, 대조군 및/또는 중재결과 측면에서 충분히 동질적이어야 한다.
6. 모든 체계적 문헌고찰은 중재결과 자료가 제시되는 내용과 방법 측면에서 충분히 동질적이어야 한다.
7. 제시된 중재결과 자료의 양과 유형은 overview의 연구 질문과 목적을 해결하기에 충분해야 한다.
8. 모든 체계적 문헌고찰은 비뚤림위험이 충분히 낮아야 한다,

26.4 Overview 수행을 위한 방법

Overview를 시행할 때는 선별과 포함 결정, 방법론적 질/비뚤림위험 평가, 자료 추출 등을 2명의 저자가 독립적으로 실시하는 것이 바람직하며, 불일치 해소를 위한 프로세스가 마련되어 있어야 한다. Overview를 수행하는 모든 방법을 사전에 고려하고 프로토콜에 자세히 설명해야 한다.

26.4.1 Overview의 중요한 방법론적 한계에 관한 참고 사항

Overview는 종종 동일한 상태 또는 인구집단에 대해 서로 다른 중재에 대한 둘 이상의 체계적 문헌고찰로부터 근거를 제시하지만, 복수 중재의 비교 효과에 대한 추론을 도출하는 경우에는 사용하지 않아야 한다. 예를 들어, 두 개의 체계적 문헌고찰을 포함하는 overview에서 체계적 문헌고찰 1은 중재 A와 중재 B를 비교하는 연구를 포함하고 있으며, A가 B보다 더 효과적이라는 것을 확인하였다. 체계적 문헌고찰 2는 중재 B와 중재 C를 비교하는 연구를 포

함하고 있으며, B가 C보다 더 효과적이라는 것을 확인하였다. Overview 저자들은 A가 C보다 더 효과적이라는 결론을 내리고 싶어 할 것이다. 그러나 overview에서는 간접 비교를 권장하지 않는다. 이는 특히 동일한 상황 또는 인구집단에 대해 여러 개의 중재를 검토하는 overview를 수행하는 저자에게 해당되며, overview에 포함된 체계적 문헌고찰이 메타 분석을 포함하건 단순한 결과 서술 요약을 제공하건 마찬가지이다. 그 이유는 간접 비교에 필요한 가정을 체계적 문헌고찰에 제공된 정보만 사용하여 평가할 수 있는 경우는 거의 없기 때문이다.

동일한 상황 또는 인구집단에 대해 여러 개의 중재를 검토하는 overview는 종종 서로 다른 체계적 문헌고찰의 자료를 비교한다. 때때로 이러한 자료는 같은 표나 그림에 나타난다. 이러한 방식으로 자료를 제시하는 overview는 의도하지 않게 독자들로 하여금 간접 비교를 하도록 권장할 수 있다. Overview가 부적절한 간접 비교를 촉진할 수 있는 경우 overview 작성자는 체계적 문헌고찰에서 '비교'를 피해야 하며, 다음과 같은 방법을 사용할 수 있다.

- 적절하게 표현된 연구 질문 및 목적을 사용(예 :'질병 X를 치료하는 데 가장 효과적인 중재는 무엇입니까?'보다는 '어떤 중재가 질병 X를 치료하는 데 효과적입니까?'라고 표현).
- 연구결과와 결론을 적절하게 해석 (예 :``위약과 비교하여 중재 A와 D는 질병 X 치료에 효과적이며 중재 B와 C는 효과적이지 않을 수 있다 '').
- 독자에게 임의적인 간접 비교와 관련된 위험에 대한 명확한 설명을 제공 (예 : 'A vs. 위약의 효과 추정치가 C vs. 위약의 효과 추정치의 2배이므로 중재 A가 중재 C보다 효과적이라고 결론을 내릴 수도 있다. 그러나 두 가지 중재를 평가하는 연구는 여러 가지 면에서 달랐으며, 우리는 독자들이 이러한 유형의 부적절한 임의적 간접 비교를 하지 않도록 강력히 권고한다. ') 자료 표와 그림에도 비슷한 경고가 제공될 수 있다.

26.4.2 연구 질문의 정의

Overview 저자는 overview의 범위를 명확하게 정의하여 시작해야 한다. Overview는 일반적으로 중재 문헌고찰보다 범위가 넓지만, 그럼에도 연구 질문은 구체적이고 잘 정의되어야 한다. Overview의 연구 질문에는 인구집단, 중재군, 대조군, 중재결과, 기간 및 세팅에 대한 명확한 설명이 포함되어야 한다. 동일한 상태 또는 인구집단에 대해 다른 중재를 검토하는 overview의 경우, overview의 주요 목표는 다음과 같은 형식으로 설명 될 수 있다: [특정 그룹, 질병 또는 문제 , 세팅] '에서 '[건강 문제]'에 대한 [중재군 혹은 비교군]의 효과를 평가하기 위해 체계적 문헌고찰을 요약.

Overview는 일반적으로 범위가 넓기 때문에 서로 다른 체계적 문헌고찰의 질문에 상당한 차이가 있을 경우 연구 질문을 제한할 필요가 있을 수 있다. 예를 들어, (복수의 질병 대신) 단일 질병 또는 특정 연구대상자 특성(예: 특정 연령 집단, 질병 중증도, 세팅 또는 동반 질환의 유형)으로 제한할 수 있다. 범위를 제한할지 여부와 방법을 결정할 때 의사 결정자의 관점을 염두에 두고 연구 질문이 임상적으로 적절하고 유용하게 유지되도록 해야 한다. 모든 제한에

는 적절한 근거가 있어야 한다.

Overview는 포함된 체계적 문헌고찰의 적격 기준에 제한을 받는다. 따라서 연구가 진행하면서 연구 질문(및 그들의 방법론)을 수정하거나 수정할 필요가 있을 수 있다. 하지만 사후 수정을 할 때 확인된 비뚤림을 피해야 하며, 모든 수정은 근거를 가지고 문서화해야 한다.

26.4.3 체계적 문헌고찰 포함 기준의 개발

앞에서 언급한 연구 질문을 통해 포함 기준을 결정하여야 한다. 포함 기준에는 모든 관련 특성(인구집단, 중재군, 비교군, 결과 측정, 기간, 세팅)에 대한 명확한 설명과 포함될 연구설계에 대한 정보(예 : 체계적 문헌고찰)가 포함되어야 한다. 제3장은 정보가 도움이 되며, 내용은 중재 문헌고찰 에 대한 것이지만 overview에도 도움이 된다.

체계적 문헌고찰을 포함할 때에도 다음의 세 가지 고려사항이 적용된다.:

첫째, 출판물이 '체계적 문헌고찰'인지를 결정하는데 사용할 기준을 명확하게 지정해야 한다. Overview 저자는 포함여부를 결정하기 위해 특정 기준을 추가해야 한다(예 : "명시적이고 재현 가능한 방법론", 포괄적 검색, 포함된 연구의 유효성을 평가하기 위한 수용 가능한 방법 등). 코크란 리뷰는 체계적 문헌고찰에 대한 코크란 정의를 준수하나 코크란 이외의 간행물은 그렇지 않을 수 있다. '체계적 문헌고찰'이라는 명칭이 붙었다고 체계적 문헌고찰의 정의를 충족하는 것은 아니며, '체계적 문헌고찰'로 분류되지 않은 일부 간행물이 체계적 문헌고찰의 정의를 충족할 수도 있다. 따라서 포함에 관한 결정을 내릴 때 사전 설정된 기준에 중점을 두어야 한다.

둘째, 무작위 임상 시험의 체계적 문헌고찰만을 포함할 것인지, 아니면 관찰 연구와 같은 여러 연구설계를 포함하는 체계적 문헌고찰을 포함할 지를 고려해야 한다. 현재는 RCT와 관찰연구의 자료를 결합하는 것을 권장하지 않는다. 따라서 비무작위 연구에 대한 체계적 문헌고찰을 포함하는 경우 자료를 별도로 제시되거나 NRS를 포함하는 체계적 문헌고찰의 자료를 사용할 수 없을 수 있다.

셋째, overview 저자는 동일한 질병에 대해 동일한 중재에 대한 둘 이상의 체계적 문헌고찰을 접할 가능성이 있다. 저자는 이러한 중복 체계적 문헌고찰을 overview에 포함시킬지 여부와 방법을 사전에 검토해야 한다.

26.4.4 중복된 체계적 문헌고찰의 관리

출판되는 체계적 문헌고찰의 수가 증가함에 따라 동일한(또는 매우 유사한) 연구 질문을 다루고 일차 연구 중 대부분(전부는 아니지만)을 포함하는 두 개 이상의 관련 체계적 문헌고찰을 확인할 수 있다. Overview에서 이러한 "중복 검토"를 포함하는 것과 관련된 두 가지 주요 방법론적 문제가 있다:

첫째, 유사한 복수의 체계적 문헌고찰을 포함하면 overview에 비뚤림을 초래할 수 있다. 중

복되는 문헌고찰을 포함하면 overview에 중재결과 자료를 이중 계상하도록 할 수 있다. 중재결과 자료를 요약하려는 경우 중재결과 자료를 이중 계산하면 일부 일차 연구의 자료가 너무 큰 영향을 미칠 수 있다. Overview 저자가 중재결과 자료를 재분석하려는 경우, 중재결과 자료를 이중 계산하여 치료 효과에 대해 지나치게 정밀한 추정치를 생성할 수 있다.

둘째, 중복되는 문헌고찰이 포함된 overview는 방법론적으로 복잡하다. Overview 저자가 체계적이고 투명한 방법으로 중복 검토 결과를 검색, 식별, 포함, 질 평가, 자료추출 및 분석, 보고하는 방법을 결정함에 따라 overview 프로세스의 모든 단계는 더 많은 시간과 자원이 소요될 것이다. 이는 중복되는 문헌고찰이 다양한 행위, 질 및 보고를 보이고, 불일치한 결과, 결론을 가지고 있는 경우에 특히 그러할 것이다.

각 일차 연구의 중재결과 자료가 중복되는 문헌고찰에서 단 한 번만 추출되도록 함으로써, 저자는 중복되는 문헌고찰에서 중재결과 자료를 이중 계산하는 것을 피할 수 있는가? 관련 체계적 문헌고찰 사이에 중복이 있을 경우 중복되는 일차 연구의 중재결과 자료를 이중 계산하지 않도록 주의할 수 있다. 이는 모든 관련 체계적 문헌고찰의 모든 중재결과 자료가 overview에 포함되도록 보장하는 유일한 방법이다. 그러나 위에서 설명한 바와 같이 이러한 결정은 시간-집약적이고 방법론적으로 복잡하다. 혹은 특정한 체계적 문헌고찰을 우선시하기 위해 사전 정의된 기준을 사용하여 이런 상황을 피할 수 있다. 저자는 중복되지 않는 모든 문헌고찰을 포함하고 가장 최신의, 질이 높고, 관련성이 높은 문헌고찰을 선택할 수 있다. 이렇게 하면 overview에 포함된 중재결과 자료의 양을 최대화하는 동시에 중복되는 문헌고찰과 관련된 잠재적인 문제를 피하게 할 수 있다.

또한 Overview에서 제대로 수행되지 않은 체계적 문헌고찰을 배제하고자 하는 저자는 위의 포함 기준 중 하나를 적용하기 전에 질/비뚤림위험 평가 결과를 배제 기준으로 사용할 수도 있다.

26.4.5 체계적 문헌고찰 검색

연구 질문의 정의, 체계적 문헌고찰 포함 기준의 개발, 중복되는 체계적 문헌고찰과 관련된 이슈를 어떻게 다룰 것인가를 검토하는 등 프로토콜을 개발했으면, 포괄적이고 재현 가능한 문헌 검색을 실시하여야 한다. Overview에서 중복되는 문헌고찰을 다루는 방법에 대한 결정을 위해 문헌 검색을 이미 수행했을 수 있다.

코크란 리뷰만 포함하려면, CDSR만 검색하면 된다. Overview에 코크란이 아닌 체계적 문헌고찰이 포함된다면 추가 데이터베이스와 체계적 문헌고찰 DB를 검색해야 한다. 일반적으로 MEDLINE/Pubmed 및 EMBASE는 대부분 체계적 문헌고찰을 색인화하고 있으며, 추가 지역 및 주제별 데이터베이스(예: LILACS, CINAHL, PsycINFO) 및 에피스토메니코스(www.epistemonikos.org).와 같은 체계적 문헌고찰 DB를 검색할 수 있다.

체계적 문헌고찰 연구설계에 특정한 검색어와 MeSH 표제어(예: ‘systematic review’, ‘meta-analysis’)를 사용하고 검증된 체계적 문헌고찰 검색 필터를 사용하여 이를 수행할 수

있다. 검증된 검색 필터 목록은 https://sites.google.com/a/york.ac.uk/issg-search-filters- resource/filters-to-identify-systematic-reviews 에서 확인할 수 있다.

26.4.6 체계적 문헌고찰 선택 기준

26.4.6.1 포함 기준을 충족하는 체계적 문헌고찰 확인

문헌 검색을 통해 찾은 논문에 대해 overview의 적격성 기준을 충족하는지 여부를 평가해야 한다. Overview에서 중복 문제를 다루는 방법에 대한 결정을 확인하기 위해 이 정보를 사용하고자 하는 경우 체계적 문헌고찰을 이미 선택했을 수 있다.

Overview에 포함하기 위해 시행하는 코크란 리뷰 평가와 관련된 두 가지 고려사항이 있다. 첫째, CDSR 검색은 프로토콜을 검색할 수 있다. 둘째, 중재 문헌고찰이 충분히 최신의 것이 아닌 경우가 있을 수 있다. 이 두 경우 모두, 관련 문헌고찰의 업데이트 시한이 가까운지 또는 업데이트되는 과정에 있는지를 해당 문헌고찰 그룹에 문의해야 한다. 만약 그렇다면, 업데이트된 중재 문헌고찰의 사전 공개 버전을 얻는 것이 가능할 수 있으며, 이에 대한 overview 포함 여부를 평가할 수 있다. 저자는 '평가 대기 중인 문헌고찰 특성'이라는 제목 아래 overview 참조 목록에 프로토콜을 포함해야 한다. 만일 코크란 리뷰가 아닌 경우 반드시 '체계적 문헌고찰'에 대한 사전 지정된 정의를 준수해야 한다.

Overview의 범위가 관련 체계적 문헌고찰의 범위보다 좁은 경우, 체계적 문헌고찰에 포함된 일차 연구의 하위군만이 overview의 적합성 기준을 충족할 수 있다. 예를 들어, Cates 등(2012년)은 소아에서 포마테롤이나 살메테롤의 안전성을 조사하는 overview를 실시했지만, 많은 관련 체계적 문헌고찰은 성인 대상 일차 연구를 포함하고 있었다. 따라서 포함된 체계적 문헌고찰 내에서, 저자들은 소아에게서 수행된 일차 연구만을 포함시켰다.

26.4.6.2 일차 연구 추가 검색 수행

때로는 포함 기준을 충족하는 모든 체계적 문헌고찰을 확인한 후에도 적용범위의 중요한 격차가 남아있는 경우가 있다(예: 포함된 체계적 문헌고찰에서 중요한 중재가 검토되지 않거나 중요한 중재에 대한 체계적 문헌고찰이 최신 것이 아닐 수 있음). 드문 경우지만, 저자들은 포함된 체계적 문헌고찰의 부족한 부분을 극복할 수 있는 일차 연구에 대한 보충 검색을 실시하는 것을 고려할 수 있다. 그러나 이러한 사항을 고려하는 저자는 동일한 overview 내에서 체계적 문헌고찰과 일차 연구를 병행할 때 관련된 추가적인 복잡성으로 인해 overview 형식의 적절성을 다시 고려해야 한다. 추가 일차 연구를 overview에 적절히 통합하는 방법에 대한 지침은 없다.

26.4.7 포함된 체계적 문헌고찰 내에서 일차 연구 중복 평가

일단 포함된 체계적 문헌고찰의 최종 목록을 작성하게 되면, 어떤 일차 연구가 체계적 문헌

고찰에 포함되는지를 파악하는 것이 중요하다.

중복의 양을 시각적으로 보여주기 위해 표 26-3과 유사한 인용 매트릭스를 작성하는 것이 유용할 수 있다. 또한 중복되는 일차 연구의 수와 크기, 그리고 분석에서 차지하는 가중치의 양을 서술적으로 설명해야 한다. 또한 체계적 문헌고찰 사이에 일차 연구 중복 정도를 수치로 측정하여 제공하는 '보정 대상 영역'(corrected covered area)을 계산하기를 원할 수 있다. Pieper et al (2014)은 인용 매트릭스 작성, 중복 설명 및 보정 대상 영역 계산에 대한 자세한 지침을 제공하였다. 포함된 체계적 문헌고찰에 다수의 중재 비교가 포함된 경우, 각 비교에 대해 일차 연구의 중복 양을 별도로 평가할 수 있다. 일차 연구의 범위와 성격에 대한 정보는 출판된 overview에 명확하게 보고되어야 하며, 특히 방법론적 또는 논리상의 이유로 일차 연구 자료의 이중 계산을 피할 수 없는 overview의 경우 더욱 그러해야 한다.

중복되는 정도를 도표화할 때 참고문헌이 동일하기 때문에 중복되는 일차 연구는 체계적 문헌고찰을 통해 쉽게 식별할 수 있다. 그러나 동일한 일차 연구의 서로 다른 특성(예: 다른 하위군, 비교군, 결과 및/또는 시점)을 설명하기 위해 서로 다른 참고문헌이 다른 체계적 문헌고찰에서 인용되는 경우 중복되는 일차 연구는 체계적 문헌고찰에서 쉽게 식별되지 않을 수 있다.

표 26-3. 포함된 체계적 문헌고찰에 포함된 주요 연구를 도식화하는 표 템플릿*

	문헌고찰 1	문헌고찰 2	문헌고찰 3	[...]	문헌고찰 X
일차 연구 1					
일차 연구 2					
일차 연구 3					
[....]					
일차 연구 X					

* 관련 셀에 'X', 'Yes', 'Included' 또는 이와 유사한 메모를 넣어 어떤 체계적 문헌고찰이 어떤 일차 연구를 포함하는지 표시

26.4.8 포함된 체계적 문헌고찰로부터 자료 수집, 분석, 제시 : 개요

Overview에 포함된 체계적 문헌고찰에서 몇 가지 유형의 자료를 추출해야 한다. 여기에는 체계적 문헌고찰(및 포함된 일차 연구)의 질/비뚤림위험, 체계적 문헌고찰(및 포함된 일차 연구)의 특성 서술, 정량적 중재결과 자료, 그리고 주요 중재결과에 대한 근거 수준 등이 포함된다. 방법론적 질/비뚤림위험 평가와 자료 추출은 두 명의 저자가 개별적으로 수행하여 불일치 해결을 위한 프로세스를 시범적으로 실시하는 것이 바람직하다(Higgins 외 2016).

체계적 문헌고찰에서 관심이 있는 자료가 누락되었거나 부적절하게 보고되었거나 다른 방

식으로 보고되었을 경우 어떻게 진행할지 미리 검토해야 한다. 때로 overview에서 적용범위의 차이를 간단히 언급하고 체계적 문헌고찰에서 특정 자료를 이용할 수 없다고 기술할 수 있다. 또는 일차 연구에서 누락된 자료를 직접 추출할 수 있다. 기반이 되는 일차 연구를 다시 언급하는 것은 overview의 포괄성과 정확성을 향상시킬 수 있지만, 추가적인 시간과 자원을 필요로 할 것이다. 저자가 일차 연구로부터 대량의 자료를 추출해야 된다는 사실을 확인할 경우, overview 형식의 적절성을 다시 고려해야 하며, 대신 체계적 문헌고찰을 실시하는 것을 고려할 수 있다.

26.4.9 포함된 체계적 문헌고찰의 방법론적 질/비뚤림위험 평가

방법론적 질은 체계적 문헌고찰 저자가 연구를 수행하고 가능한 한 가장 높은 수준으로 보고한 범위에 대한 비평적 평가를 말한다. 비뚤림은 참값으로부터의 결과나 추론의 체계적 차이를 말한다. 이러한 비뚤림은 연구설계, 시행, 분석 및/또는 보고의 결함으로 인해 발생할 수 있다. 연구에 결함이 있더라도 추정치에 비뚤림이 있는지 여부를 항상 알 수 있는 것은 아니다. 더 나아가서는 정량화하기도 어렵고 때로는 비뚤림의 방향을 예측하기도 어렵다. 이러한 이유로 평가자는 '비뚤림위험'을 판단해야 한다. 저자가 overview에서 중복되는 체계적 문헌고찰을 다루는 방법에 대한 의사결정 과정을 확인하기 위해 이 정보를 사용하고자 한다면 포함된 체계적 문헌고찰의 방법론적 질/비뚤림위험을 이미 평가했을 수 있다.

Overview 저자는 세 가지 도구 중 하나를 사용하여 overview에 포함된 체계적 문헌고찰의 방법론적 질 또는 비뚤림위험을 평가할 수 있다.

AMSTAR 도구(Shea 외 2007)는 무작위배정 임상시험 체계적 문헌고찰의 방법론적 질을 평가하도록 개발되었으며, 현재까지 overview에서 가장 많이 사용되는 도구이다. 이 도구를 사용하고자 하는 연구자들은 overview에서 AMSTAR 사용에 대한 경험적 근거와 권장사항을 Pollock 외 (2017b)에서 참고할 수 있다.

AMSTAR 2 도구(Shea 외 2017)는 AMSTAR 도구의 업데이트된 버전이다. 의료 중재에 대한 무작위배정 임상시험과 비 무작위 연구를 모두 포함하는 체계적 문헌고찰의 방법론적 질을 평가하는 데 사용할 수 있다. AMSTAR2는 질 높은 체계적 문헌고찰을 식별하는 것이 목적이며, 다음과 같은 영역을 포함하고 있다: 프로토콜 등록, 문헌 검색의 적절성; 배제 연구에 대한 정당화; 포함 연구에 대한 비뚤림위험; 메타 분석 방법의 적절성; 연구결과를 해석할 때 비뚤림위험에 대한 고려; 출판 비뚤림의 존재와 영향. 이 도구는 향후 overview에서 사용하는 것이 적합할 수 있다.

마지막으로, 최근에 개발된 ROBIS 도구(Whiting 외 2016)는 overview에서 체계적 문헌고찰의 비뚤림위험을 평가하고자 하는 경우 사용할 수 있다. ROBIS는 중재, 진단, 예후 및 병인과 관련된 질문을 다루는 체계적 문헌고찰에 사용되도록 설계되었다. 이 도구는 1) 적절성 평가(선택사항으로 간주되지만 체계적 문헌고찰을 선택하는데 도움이 될 수 있음), 2) 체계적 문

헌고찰 과정의 우려 사항 확인, 3) 체계적 문헌고찰에 대한 전반적인 비뚤림위험 판단(낮음, 높음, 불확실)의 세 단계로 이루어져 있다. 두 번째 단계는 체계적 문헌고찰 과정에서 비뚤림의 원인이 될 수 있는 네 가지 영역을 포함하는데 그것은 연구 적격성 기준, 연구의 식별 및 선택, 자료 수집 및 비뚤림위험 평가, 합성 등이다. 이 도구는 ROBIS 웹 사이트(http://www.bristol.ac.uk/population-health- sciences/projects/robis/)에서 제공된다. 이 웹 사이트에는 사전 형식의 자료 추출 양식 및 자료 제시 표도 포함되어 있다.

우리는 현재 이 주제에 대한 경험적 근거가 부족하기 때문에 한 도구를 다른 도구보다 더 추천할 수 없다. 그러나 어떤 도구를 사용하든 간에, 도구의 각 질문에 대해 각 체계적 문헌고찰이 어떻게 평가되었는지, 평가의 근거 및 각 체계적 문헌고찰에 대한 전반적인 평가를 제공하는 표(해당되는 경우)를 포함해야 한다. 그런 다음, 저자는 비뚤림위험 평가 결과를 사용하여 overview의 근거를 맥락화해야 한다(예: SR 방법이 overview의 이해도와 결과에 영향을 미쳤는지 여부와 그 정도를 평가함).

26.4.10 체계적 문헌고찰에 포함된 일차 연구의 비뚤림위험에 대한 자료 수집 및 제시

Overview를 수행할 때 포함된 체계적 문헌고찰에 포함된 관련 일차 연구에 대한 영역별 혹은 전반적인 비뚤림위험 평가결과를 추출하고 보고해야 한다. 주요 비뚤림위험 영역은 무작위 배정 과정에서 발생하는 비뚤림, 의도한 중재에서 이탈로 인한 비뚤림, 중재결과 결측에 인한 비뚤림, 중재결과 측정의 비뚤림, 보고된 연구결과의 선택에 대한 비뚤림 등이 있다.

이상적으로는 포함된 각 체계적 문헌고찰에서 제시된 평가내용을 추출해야 한다. 그런 다음 평가 결과에 대한 요약이나 표를 제시할 수 있다. 단, 체계적 문헌고찰별로 다른 도구를 사용하여 비뚤림위험을 평가했을 가능성이 있다. 이러한 상황에서 상이한 비뚤림위험 결과를 최대한 추출해야 한다. 그런 다음 저자들은 두 가지 선택사항을 가지는데 평가에 대한 서술적 요약 및/또는 표로 제시된 요약내용을 제공하거나 일차 연구를 참고하여 비뚤림 영역의 누락된 부분에 대한 자료를 추출하여 기존 평가를 보완하는 것이다.

26.4.11 포함된 체계적 문헌고찰(및 일차 연구)의 서술적 특성에 대한 자료 수집 및 제시

Overview에 포함된 각 체계적 문헌고찰의 서술적 특성에 대한 정보를 추출해야 한다. 각 체계적 문헌고찰의 출발점으로, 표 26-4에 열거된 정보를 추출하는 것이 유용할 수 있다. 이 정보는 '포함된 문헌고찰의 특성' 표에 보고될 수 있다. Overview의 특정 요구사항 또는 목표에 따라 추가 설명 자료를 추출해야 할 수 있다. 또한 체계적 문헌고찰에 포함된 결과와 overview에 미리 명시된 중재결과 사이의 불일치된 내용을 본문에 기록해야 한다.

표 26-4. Overview 저자가 포함된 체계적 문헌고찰에서 추출하고자 할 수 있는 체계적 문헌고찰(및 그 일차 연구)의 서술적 특성

- 체계적 문헌고찰에 대한 기본 정보(예: 제목, 저자, 출판연도, 현재성이 있다고 마지막으로 평가한 날짜, 체계적 문헌고찰에 포함된 연구 및 연구대상자 수)
- 일차 연구에 대한 기본 정보(예: 저자, 출판연도, 연구설계, 출판 국가)
- 체계적 문헌고찰의 검색 전략(예: 검색된 데이터베이스 수, 검색된 데이터베이스명, 데이터베이스 검색일 범위, 마지막 검색 업데이트 날짜)
- 체계적 문헌고의 연구대상자(예: 연령, 성별, 인종, 질병 중증도, 동반질병, 질병의 정의, 연구 세팅 등 연구대상자 특성)
- 체계적 문헌고찰 중재(예: 중재 유형, 양, 강도, 빈도, 기간)
- 체계적 문헌고찰 비교 중재(예: 비교 중재 유형, 양, 강도, 빈도, 기간)
- 1차 및 2차 중재결과(체계적 문헌고찰 방법 파트에서 자세히 설명)
- 추가 정보(예: overview 저자의 의견, 체계적 문헌고찰 제한점, 비뚤림위험)

26.4.12 정량적 중재결과 자료 수집, 분석 및 제시

Overview에서 중재결과 자료를 분석하는 주요한 두 가지 방법이 있다. 첫 번째 방법은 중재결과 자료의 요약부분에 포함된 체계적 문헌고찰에서 제시된 것과 똑같은 자료를 제시하는 것이다. 두 번째 방법은 중재결과 자료 재분석을 통해 일부 자료는 포함된 체계적 문헌고찰에서 중재결과 자료를 추출하고, 일부는 체계적 문헌고찰에서 수행된 분석과 다른 방식으로 자료를 분석하며, 재분석된 자료를 overview에 제시하는 것이다. 가장 적절한 자료 분석 방법은 overview의 목적, 특정 주제 영역 및 포함된 체계적 문헌고찰의 특성에 따라 달라질 수 있다.

26.4.12.1 중재결과 자료의 요약

중재결과 자료를 요약하면 포함된 각 체계적 문헌고찰에 포함된 자료의 개별 서술 요약자료(효과 추정치 및 95% 신뢰 구간 포함)를 제시하여 이용 가능한 근거를 독자들에게 제공한다. 관련 체계적 문헌고찰(및 그 중재결과 자료)의 대상군 별 연구결과를 기술하고 요약하여 독자에게 체계적 문헌고찰의 내용과 연구결과를 제시한다. 이러한 방법의 목적은 관련 체계적 문헌고찰 중 중재, 대조 중재, 중재결과 혹은 연구결과를 확인하여 설명하는 것이다.

중재결과 자료를 요약할 때는 체계적 문헌고찰에서 보고된 대로 자료를 추출한 다음, 필요에 따라 텍스트, 표, 그림으로 적절히 재구성하여 제시하면 된다. 효과 추정치, 95% 신뢰 구간 및 이질성 측정치(연구가 합성된 경우)는 모두 추출해야 한다. 가능한 한 포함된 체계적 문헌고찰에 보고된 분석에 의존해야 한다.

26.4.12.2 중재결과 자료의 재분석

중재결과 자료 재분석은 포함된 체계적 문헌고찰에서 관련 중재결과 자료를 추출하고 체계

적 문헌고찰에서 수행한 원래 분석방법과는 다른 방식으로 자료(예: 메타 분석 사용)를 재분석하는 것이다. 몇 가지 이유로 중재결과 자료를 재분석할 수 있다. 첫째, overview의 목적이 다른 임상 질문에 대한 답변인 경우, 저자는 해당 질문에 대한 특정 자료(예: 성인이 아닌 소아에서의 중재 효과)만 선택하여 재분석할 수 있다. 둘째, 체계적 문헌고찰의 대부분이 특정 인구집단이나 하위군을 분석한 경우, 체계적 문헌고찰 주제 전반에 걸쳐 일관성 있는 정보가 보고되도록 나머지 체계적 문헌고찰에 이러한 분석을 적용할 수 있다. 셋째, 포함된 체계적 문헌고찰에서 다른 요약 측정값이나 모델을 사용하는 경우 자료를 재분석할 수 있다. 이는 체계적 문헌고찰 주제전반에 걸쳐 일관된 방식으로 결과를 제시할 수 있기 때문이다(예: 모든 추정치를 상대 측정치 또는 절대 추정치로 제시). 마지막으로, overview 저자는 체계적 문헌고찰에서 이전에 메타 분석되지 않았던 자료를 분석하도록 선택할 수 있다. 체계적 문헌고찰 저자들이 개별 연구에 대한 심도 있는 지식을 바탕으로 해당 분석을 시행했을 가능성이 높기 때문에 이 마지막 두 가지 경우에는 주의를 기울여야 한다. Overview 저자는 자신이 원하는 중재결과 자료를 재분석하는 방법이 적절한지 여부를 판단할 때 체계적 문헌고찰 저자가 분석 방법을 선택한 이유를 이해해야 한다.

중재결과 자료를 다시 분석하는 경우 표준 메타 분석 원칙을 사용해야 한다. 중재결과 자료를 재분석하고자 하는 경우, 포함된 체계적 문헌고찰의 임상 매개변수와 통계적 측면이 충분히 보고된 경우에만 그렇게 할 수 있다는 점에 유의해야 한다. 이러한 유형의 분석을 수행할 때 저자는 가능한 한 표준화된 방법으로 재분석한 연구결과를 제시하도록 노력해야 한다(예: 고정 또는 변량 효과 모형 사용 및 각 중재결과에 대해 일관된 효과 측정치 사용). 또한 overview 저자는 두 개 이상의 중재의 비교 효과에 대해 부적절한 비공식 간접 비교를 하지 않도록 주의해야 한다(4.1절 참조). CDSR에 접근할 수 있는 저자는 코크란 중재 문헌고찰에 대한 Review Manager 파일을 다운로드하여 자료 추출 속도를 높일 수 있다.

26.4.12.3 중재결과 자료 제시

Overview 저자는 요약되거나 재분석된 중재결과 자료를 서술적으로 또는 연구결과 표를 통해 제시할 수 있다. 표에 대한 구체적인 형식은 없지만, 중재결과 자료를 표시하기 위한 원칙을 따라야 한다. Overview 저자는 다음과 같이 할 수 있다:

- 체계적 문헌고찰에 포함된 중재결과 자료에 대한 해당 표를 포함하거나 혹은 포함하지 않고 서술적인 요약내용을 제시한다. 예를 들어, overview 저자는 체계적 문헌고찰 연구결과를 차례로 제시 하거나 중재결과 측정값을 그룹화 하여 함께 제시할 수 있다.
- 연구결과를 범주화(예: '임상적으로 중요함'또는 '임상적으로 중요하지 않음' 또는 '유효한 중재', '유망한 중재', '비효과적인 중재', '비효과적인 중재' '결론 내리기 어려움')하고, 가용한 근거를 제시함.
- 새로운 개념 틀을 제시하거나 기존 틀을 수정하여 제시한다. 예를 들어, 저자들은 중재 vs. 중재결과의 기준을 제시할 수 있다. 그런 다음 얼마나 많은 일차 연구들이 중재결과에 기여하

는지, 그리고 각 중재결과에 대한 효과의 방향을 나타낼 수 있다. 또한 포함된 체계적 문헌고찰 중재의 분류에 다른 틀을 제시할 수 있으며, 각 중재 범주의 효과를 설명할 수 있다. 중재결과 자료를 제시하는데 사용되는 틀은 프로토콜 단계에서 우선적으로 지정하거나 보고서에서 사후에 표시하여야 한다.

표 26-5는 저자가 사용할 수 있는 '결과 요약표' 템플릿이다. 결과 요약표를 작성할 때, 통계적으로 유의한 결과만 선별적으로 보고하지 않도록 주의해야 한다. 또한 다양한 체계적 문헌고찰의 자료를 단일 표 또는 그림으로 요약하기로 하였다면, 독자들에게 비공식적인 간접 비교를 하도록 할 수 있으며, 이러한 종류의 표는 체계적 문헌고찰 전반에 걸쳐 '비교'를 피하고 연구결과를 적절하게 해석하며, 독자들에게 절차를 따르도록 기술한 경우에만 사용하도록 한다.

표 26-5. '결과 요약표' 템플릿

중재 for [상태] in [인구집단]							
중재결과	중재군과 비교군	Illustrative comparative risks(95% CI)		비교 효과(95% CI)	연구 대상자수 (연구수)	근거의 확실성 (GRADE)	비고
		Assumed risk 비교군	Correspon - ding risk 중재군				
중재결과 #1							
	중재군/ 비교군 1						
	중재군/ 비교군 2						
	[...]						
	중재군/ 비교군 X						
중재결과 #2							
	중재군/ 비교군 1						
	중재군/ 비교군 2						
	[...]						
	중재군/ 비교군 X						
중재결과 'X'							
	중재군/ 비교군 1						
	중재군/ 비교군 2						
	[...]						
	중재군/ 비교군 X						

26.4.13 GRADE 도구를 사용하여 정량적 중재결과 자료의 근거 수준 평가

중재 문헌고찰과 마찬가지로, overview는 GRADE 도구를 사용하여 overview에서 사전 정의된 임상적으로 중요한 각 관심 중재결과에 대한 근거 수준(즉, 효과 추정치에 대한 신뢰도)을 평가하고 보고하도록 한다. 가능한 경우 포함된 체계적 문헌고찰에 제시된 GRADE 평가내용을 추출하여 보고해야 한다. 그러나 특히 비코크란 체계적 문헌고찰이 overview에 포함된 경우 주의해야 할 사항이 있다. 예를 들어, 일부 체계적 문헌고찰은 GRADE 평가를 포함하지 않을 수도 있고, 제한적인 GRADE 평가를 포함할 수도 있고,(개별이 아닌) 집계된 평가결과를 제시할 수도 있고, 근거 수준을 평가하기 위해 GRADE 이외의 도구를 사용할 수도 있다. 또한 overview가 체계적 문헌고찰에서 중재결과 자료를 추출하고 재분석하는 경우, 체계적 문헌고찰의 기존 GRADE 평가는 더 이상 관련이 없을 수 있다. 이 경우, 체계적 문헌고찰에 보고된 정보를 사용하여 스스로 GRADE 평가를 수행할 필요가 있는지 결정해야 한다.

26.5 참고문헌과 읽을거리

- Higgins JPT, Lasserson T, Chandler J, Tovey D, Churchill R. Standards for the conduct of new intervention Reviews. Methodological Expectations of Cochrane Intervention Reviews 2016.
- Pieper D, Antoine S, Mathes T, Neugebauer E, Eikermann M. Systematic review finds overlapping reviews were not mentioned in every other overview. Journal of Clinical Epidemiology 2014;67: 7.
- Pollock M, Fernandes RM, Becker LA, Pieper D, Hartling L. Chapter V: Overviews of Reviews. Draft version (8 October 2018) for inclusion in: Higgins JPT, Thomas J, Chandler J, Cumpston MS, Li T, Page MJ, Welch V (editors). Cochrane Handbook for Systematic Reviews of Interventions. London: Cochrane.
- Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, Moher D, Tugwell P, Welch V, Kristjansson E, Henry DA. 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.
- Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, Moher D, Tugwell P, Welch V, Kristjansson E, Henry DA. 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
- Whiting P, Savović J, Higgins J, Caldwell D, Reeves B, Shea B, Davies P, Kleijnen J,

Churchill R. ROBIS: a new tool to assess risk of bias in systematic reviews was developed. Journal of Clinical Epidemiology 2016;69: 9.

제27장 환자보고 중재결과 (Patient-reported outcomes)

핵심 포인트

- 환자보고 중재결과(Patient-reported outcomes, PRO)에 대한 요약자료는 보건의료 의사결정자가 환자에게 가장 의미 있는 중재결과에 대해 정보를 주기 때문에 중요하다고 할 수 있다.
- PRO를 포함하는 체계적 문헌고찰의 저자는 측정하려는 구성(construct), 신뢰도, 타당도 및 반응도(responsiveness)들을 포함하여 환자보고 중재결과 측정치(Patient-reported outcomes measures, PROM)가 어떻게 개발되는지 잘 이해해야 한다.
- 저자는 관심 있는 중재결과를 측정하기 위해 프로토콜 단계에서 PROM의 계층(hierarchy)을 미리 정해야 한다.

27.1 PRO 소개

27.1.1 PRO란 무엇인가?

환자보고 중재결과(Patient-reported outcomes, PRO)는 "환자가 자신의 건강상태를 임상의사 등의 해석 없이 직접적으로 보고하는 것"이다. PRO는 임상시험과 비무작위 연구에서 사용되는 바이오마커, 이환율의 측정(예. 뇌졸중, 심근경색), 질병부담(예. 입원), 생존율 등 중재결과를 보완하는 평가 방법 중 하나이다.

환자보고 중재결과 측정도구(Patient-reported outcome measures, PROM)는 PRO를 측정하는데 사용되는 도구이며 대부분 자가-보고 설문지이다. 비록 연구자는 대리(proxy) 지표나, 보호자, 보건의료 전문가 또는 부모 및 보호자의 관찰을 통해 환자 관련 중재결과를 설명할 수는 있지만, 이는 PROM이 아니라 임상의가 보고하거나 관찰자가 보고한 중재결과이다.

PROM을 포함하는 연구에 대한 체계적 문헌고찰과 메타분석을 수행하고 그 연구결과를 해석하는 것은 간단하지 않다. 이 지침은 그러한 어려움에 다소나마 도움이 될 수 있을 것이다.

이 장의 목적은 다음과 같다: 1) PRO로 알려진 중재결과의 범주와 보건의료 의사결정자를 위한 중요성을 설명한다. 2) PRO를 포함할 때 체계적 문헌고찰 저자가 고려해야 할 신뢰도, 타당도 및 반응도(responsiveness)에 관련된 주요 문제를 보여준다. 3) PRO의 구조 및 내용(영역, 문항)을 다루고 다른 PRO의 정보를 결합하기 위한 지침을 제공한다. 이 장의 초점은 PRO를 사

용하는 것, 그리고 메타분석에 포함시킬 PRO를 선택할 때 중요한 것이 무엇인가이다.

27.1.2 왜 PRO인가?

PRO는 치료의 이득과 위해에 대한 환자의 시각을 제공하고 치료의 이득과 위해를 직접 측정하며, 사망률, 이환율, 바이오마커 등 다른 중재결과를 넘어서는 정보를 제공하고, 환자와 가족들에게는 가장 중요한 중재결과이기도 하다. 자가-보고 연구결과는 성과-관련 중재결과(performance-related outcomes)나 임상의가 보고한 중재결과, 혹은 바이오마커와 같은 생리학적 중재결과와 관련성이 낮은 경우가 있다. COPD에서 보고된 FEV1과 QoL 간의 보고된 연관성은 약하며, 골관절염에서 영상 소견과 환자보고 통증 사이에 불일치가 있다.

PRO는 외부에서 관찰 가능한 환자 중요 중재결과(patient-important outcomes)가 드물거나 사용할 수 없을 때 필수적이다. 통증 증후군, 피로, 과민성 장 증후군, 성기능 장애, 정서적 기능 장애, 생리학적 측정이 제한적이거나 사용할 수 없는 메스꺼움 및 불안과 같은 부작용을 비롯한 여러 가지 상태의 치료 영향을 평가하는 중재결과로 PRO는 합리적인 전략 중 유일하다.

27.2 문헌고찰을 구성하기

이 절에서는 PROM에 대해 자세히 설명하고 문헌고찰에서 PROM을 포함할 때 고려해야 할 몇 가지 문제를 논의한다.

건강 상태 측정 문헌에서 사용되는 일반적인 용어는 구성(construct)이다. 구성이란 PROM이 측정하려고 하는 것, 통증, 신체 기능 또는 우울 기분과 같은 PROM을 정의하는 개념을 나타낸다. 구성은 연구자들이 PROM을 통해 포착하기를 원하는 속성(attribute)이다.

PROM과 그 도구의 구성을 분류하고 명칭을 부여하는 여러 방법이 있다. 예를 들어 징후 같이 환자로부터 보고 받은 것(어떤 상태의 관찰 가능한 표현), 감각(질병 또는 치료에 기인할 수 있는 증상), 행동과 능력(기능 상태), 일반적인 인식이나 웰빙의 느낌, 일반적인 건강, 치료 만족도, 부작용 보고, 치료 순응도, 그리고 사회 또는 지역 사회 행사 참여, 건강관련 삶의 질 등이 그런 것이다.

연구자는 인터뷰, 자가-보고 설문지, 일지, 소형 디바이스 또는 컴퓨터 같은 다양한 인터페이스를 통하는 등 다양한 접근법을 사용하여 환자의 관점을 포착한다. 문헌고찰 저자는 환자에게 중요한 구성을 확인하고, 임상시험에서 보고된 PROM이 이러한 구성, 다루는 범위, 사용된 PROM의 특성을 어느 정도 다루었는지를 판단한 후 이 정보를 독자들에게 전달해야 한다.

중요한 PRO인 HRQoL에 초점을 맞추면서, 건강 관련 환자 경험(자가-치료 및 신체적, 정서적 사회적 기능 등)을 모두 커버하려는 몇몇 시도가 있다. 이러한 접근법을 일반적 도구(generic instruments)라고 부른다. 여기에는 EuroQol 5 dimensions question-

naire(EQ-5D), Health Utilities Index(HUI)와 같은 효용(utility) 측정 방법도 있고, Short Form 36-item(SF-36) 또는 SF-12 와 같은 건강 프로파일도 있다. 이들 도구들이 건강 프로파일 분야에서 가장 많이 이용되고 있다. PRO를 측정하기 위한 또 다른 접근법으로 훨씬 구체적인 구성(construct)에 초점을 맞추는 것이 있다. 즉 기능(예. 수면, 성기능), 질병(예. 천식, 심장마비), 인구집단(예. 허약한 노인), 증상(통증, 피로)에 따라 특화된 도구들이다.

흔히 '삶의 질', '건강 상태', '기능 상태', 'HRQoL' 및 '웰빙'이라는 용어를 느슨하게 상호 교환하여 사용하기도 한다. 따라서 체계적 문헌고찰 저자는 PROM이 실제로 측정한 구성을 신중하게 고려해야 한다. 그렇게 하려면 PROM에 포함된 항목이나 질문을 검토해야 한다.

고려해야 할 또 다른 문제 중 하나는 개별 도구 항목에 가중치가 적용되는지 여부와 방법이다. 많은 접근법이 가중치를 적용하고 있는데 경제성 평가를 위해 고안된 효용 도구는 항목 가중치 부여에 대해 더욱 중점을 두어 궁극적으로 HRQoL을 죽음과 완전한 건강 사이의 연속 상태로 제시하려고 시도한다. 많은 PROM은 전체 점수 계산에서 항목별로 동등한 가중치를 주는데, 합리적인 접근법이다.

표 27-1은 임상시험에서 PROM의 구조 및 측정방법을 포함하여 PROM을 고려하여 보고하기 위한 틀을 제시한다. 문헌고찰을 할 때 포함된 연구에서 사용된 PROM을 잘 이해하면, 연구 결과의 적절한 분석과 근거 수준 평가에 많은 도움을 받는다.

표 27-1. 임상시험에서 PROM을 설명하고 평가하기 위한 체크리스트(Guyatt et al, 1997)

1. PROM이 평가하고 있는 것은 무엇인가?
1.1. 연구에서 PROM이 평가하고 있는 개념이나 구성은 무엇인가?
1.2. (있다면) 개념이나 구성을 선택하는데 어떤 이유가 있는지 저자가 제시하였는가?
1.3. 환자들이 PROM을 개발하는데 참여하였는가(예. 포커스 그룹 인터뷰, 설문조사 등)?
2. 보고 누락
2.1. 환자의 건강(예. 증상, 기능, 자각)이나 삶의 질(예. 전반적인 평가, 삶의 만족도)에서 중요한 부분이 이 연구에서 보고되지 않았는가? 해당 상태에 대해 '핵심 중재결과 세트'에서 찾아보는 것이 도움이 될 것이다.
3. 측정 전략은 무엇인가?
3.1. 연구자가 사용한 도구는 단일 지표 생산 도구, 지수 번호(index number) 생산 도구, 프로파일, 도구 배터리 생산 도구 중 무엇인가?
3.2. 연구자는 특이적 측정도구를 사용했는가? 혹은 일반적 측정도구를 사용했는가? 혹은 둘 다인가?
4. 도구가 작동해야 하는 방식으로 작동하였는가? - 타당도
4.1. 현재 인구집단에서 사용하기 위해 사전에 타당도의 근거가 제시되었는가?
5. 도구가 작동해야 하는 방식으로 작동하였는가? - 반응도

5.1. PROM은 환자 상태의 변화가 작더라도 중요한 변화를 감지할 수 있는가?
6. (할 수 있다면) 효과의 크기를 독자들이 이해할 수 있도록 할 수 있는가? - 해석능력
6.1. 중재가 PROM에 명백한 영향을 미친 경우, 사용자에게 그 효과가 사소한 것인지, 작지만 중요한 것인지, 중등도인지 또는 큰 지에 대한 감각을 제공할 수 있는가?

27.3 근거의 평가(appraisal)

27.3.1 PRO의 측정: 단일 vs. 다중 시점

유용한 도구는 관심 있는 여러 상황을 구별할 수 있어야 한다. 어떤 연구결과를 단지 한 시점에만, 사용할 수 있는 경우 PROM의 핵심 문제는 점수가 더 바람직한(desirable) 점수를 가지고 있는 사람들을 덜 바람직한 점수를 가진 사람들과 구별할 수 있어야 한다는 점이다. 이러한 맥락에서 핵심 측정 이슈는 신뢰도과 횡단 구성 타당도(cross-sectional construct validity)이다.

무작위배정 임상시험과 같은 종단 연구에서 연구자는 보통 임상시험의 시작과 중재 후처럼 여러 시점에서 측정치를 얻는다. 이러한 맥락에서 PROM은 부정적인 변화를 경험하는 사람, 덜 긍정적인 변화를 경험하는 사람 혹은 변화가 없는 사람과 긍정적인 변화를 경험하는 사람을 구별할 수 있어야 하며, 그 변화의 크기를 정확하게 평가할 수 있어야 한다. 이 경우 핵심 측정 이슈는 반응도과 종단 구성 타당도이다.

27.3.2 신뢰도(reliability)

많은 사람들이 신뢰도를 점수의 안정성 즉 응답자에게 도구를 반복적으로 제시할 때 동일한 점수를 얻는 것만으로 생각한다. 안정성(즉 측정 오류의 문제)이 중요 요소이기는 하지만 충분하지는 않다. 만족스러운 도구는 측정 오류가 있더라도 개인들을 구별할 수 있어야 한다.

따라서 신뢰도(reliability statistics)는 응답자 간의 변동성(일반적으로 신뢰도의 분자)과 전체 가변성(응답자 간의 변동성과 응답자 내의 변동성)의 비율을 조사한다. 신뢰도를 측정하기 위해 가장 일반적으로 사용되는 통계는 범주형 자료의 카파 계수, 순서형 자료의 가중치 카파 계수 및 연속형 자료의 급내 상관 계수(intraclass correlation coefficient, ICC) 이다.

어떤 무작위배정 임상시험에서 실험 중재가 비교 중재보다 더 우월하다는 것을 보여주지 못했을 때 문헌고찰에서 최대 관심사는 신뢰도의 한계가 될 것이다. 그 이유는, 신뢰도가 부족할 때 존재하지 않는 중재의 효과를 만들 수는 없지만, 무작위 오류로 인해 정확한 중재 효과를 검출하지 못하게 할 수 있기 때문이다. 체계적 문헌고찰에서 중재가 PROM에 영향을 준다

는 근거를 찾지 못하면 문헌고찰 저자는 효과가 부족해서 보다는 신뢰도가 낮아서 그런 것인지 고려해야 한다(예. 신뢰도 계수가 0.7 미만인 경우).

27.3.3 타당도(validity)

타당도는 어떤 도구가 측정하고자 하는 바를 측정하는지 여부와 관련이 있다. 내용 타당도(content validity)는 대개 환자 및 임상의사가 측정해야할 것을 실제로 측정하고 있는 정도를 관련성과 포괄성 측면에서 평가하는 것이다.

구성 타당도(construct validity)는 평가 척도 사이에 존재해야 하는 논리적인 관계를 조사하는 것과 관련이 있다. 예를 들어 COPD 환자에서 트레드밀 운동 능력이 낮은 환자는 일반적으로 운동 능력이 높은 환자보다 일상생활에서 더 많은 호흡곤란(호흡의 부족)을 갖게 될 것으로 예상한다.

시간 경과에 따른 변화를 평가하는데 관심이 있는 경우, 즉 중재 전과 후의 측정이 가능한 경우 평가할 때, 변화 점수의 상관관계를 조사한다. 예를 들어, 러닝머신 운동 능력이 약화된 COPD 환자는 일반적으로 호흡곤란이 심해지는 반면, 운동 능력이 호전되는 환자는 호흡곤란이 적어져야 한다. 유사하게, 새로운 감정 기능을 평가하는 도구는 기존의 감정 기능 측정에서 좋다고 하는 경우 동시에 개선의 효과를 보여주어야 한다. 이런 과정을 종단 구성 타당도를 검증하는 것이라고 한다. 문헌고찰 저자는 임상연구에서 사용된 PROM의 타당도에 대한 근거를 찾아야 한다. 불행하게도, PROM을 사용한 무작위배정 임상시험의 보고에서 사용하는 도구의 타당도에 대한 근거를 거의 제시하지 않지만, 가능하다면 설문지가 예전에 검증되었는지 찾아보고, PROM에 대해 추가적으로 출판된 정보를 찾아보는 것이 좋다. 이상적으로는, 문헌고찰 저자는 해당 도구의 측정 속성에 대한 체계적 문헌고찰을 찾아봐야 한다. Consensus-based standards for the selection of health measurement instruments(COSMIN) 웹사이트는 이러한 문헌고찰의 데이터베이스를 제공한다(COSMIN Database of Systematic Review). 또한 Patient-Reported Outcomes and Quality of Life Instruments Database(PROQOLID)는 1,000개가 넘는 PRO에 대해 측정 속성에 대한 문서를 제공한다.

체계적 문헌고찰에서 사용된 PROM의 타당도가 명확하지 않은 경우 문헌고찰 저자는 PROM이 문헌고찰의 계획된 중재결과에 적절한 척도인지, 아니면 제외되어야 하는지 고려해야 하고(이상적으로는 프로토콜 단계에서 고려해야 함), 포함된 연구결과는 주의해서 해석되어야 한다. 예를 들어, 치질에 대한 플라보노이드의 문헌고찰에서, 임상시험 저자는 PROM을 사용하여 통증과 출혈에 대한 환자의 경험을 확인하였는데, 이들 도구에 대한 검증이 없으면 이러한 도구가 중재와 비교 중재를 구분할 수 있는 의미 있는 자료를 제공할 수 있는지에 대해 우려가 제기될 것이다.

타당도에 대한 마지막 관심사는 측정 도구가 다른 인구집단이나 문화적으로나 언어적으로 다른 환경에서 사용된 경우에 발생한다. 이상적으로는, PROM은 각 연구에서 재검증되어야

하지만, 체계적 문헌고찰의 저자는 이 근거만으로 너무 비판적이지 않도록 주의해야 한다.

27.3.4 반응도(responsiveness)

무작위배정 임상시험 연구대상자에 대한 측정은 일반적으로 중재 전과 후에 가능하다. 따라서 PROM은 임상시험 전반에서 동일하게 유지되거나 나아지거나 악화되는 환자를 구별할 수 있어야 한다. 이러한 측정 특성을 반응도라고 부른다. 다른 말로 변화에 대한 민감성 혹은 변화를 감지할 수 있는 능력이다.

신뢰도와 마찬가지로, 메타분석에서 중재군과 대조군의 차이에 대한 근거가 없을 때, 반응도가 문제가 된다. 변화를 측정할 수 있는 능력이 떨어지는 도구는 중재가 환자의 감정을 향상시키는 경우에도 도구가 그 개선을 감지하지 못하면, 위음성의 결과가 나타날 수 있다. 이 문제는 특히 일반 도구에서 문제가 되는데 그 이유는 일반도구가 HRQoL의 모든 관련 영역을 포괄하는 장점이 있지만 특정 사용 상황에 필요한 세부 정보가 없을 수 있기 때문이다. 따라서 어떤 연구에서 중재군과 대조군 사이에 PROM에 차이가 없다면 도구의 반응도가 부족한 것도 가능한 이유가 된다. 문헌고찰 저자는 반응도가 입증된 출판된 근거를 찾아야 한다. 반응도에 대한 선행 근거가 없다면 이는 무작위배정 임상시험에서 얻어진 근거에 대해 덜 확신하게 만드는 잠재적인 이유가 된다.

27.3.5 보고 비뚤림(reporting bias)

PRO에 초점을 맞춘 연구는 종종 동일하거나 유사한 구성을 측정하기 위해 여러 가지 PROM을 사용한다. 이러한 상황은 선택적으로 연구결과를 보고하는 비뚤림위험이 생길 수 있는데, 이는 임상시험 저자가 임상시험의 연구결과에 기초하여, 즉 더 큰 중재 효과나 통계적으로 의미 있는 p 값을 제시하기 위해 PROM 중 일부분만 고르는 것을 말한다.

PRO에 초점을 맞춘 체계적 문헌고찰은 이 문제에 대해 경계해야 한다. 소수의 적격 연구만이 특정 PROM의 연구결과를 보고한 경우, 특히 PROM이 연구 프로토콜 또는 연구방법 부분에 언급된 경우, 또는 일반적으로 기대되는 핵심적인 중재결과인 경우, 문헌고찰 저자는 보고 비뚤림의 가능성을 기록하고 GRADE 평가에서 근거 수준을 낮게 평가하는 것을 고려해야 한다.

27.4 합성 및 근거의 해석

27.4.1 여러 PROM 중에서 선택하기

특정 PRO의 정의는 연구마다 다를 수 있으며 이는 서로 다른 도구(즉, 다른 PROM)의 사용을 정당화 할 수 있다. 정의가 비슷하더라도 연구자는 PRO를 측정하기 위해 여러 가지 다른

도구를 선택할 수 있는데, 특히 어떤 도구를 사용할지에 대한 합의가 부족한 경우 그렇다.

임상시험에서 둘 이상의 도구에 대한 결과를 보고할 때 - 이상적으로는 프로토콜 개발단계에서 연구결과를 알지 못한 상태에서 - 보고된 측정 속성을 기반으로 한 계층 구조를 만들어야 하는데 , 각 PROM이 무엇을 측정하는지 자세히 이해해야 하고, 신뢰도, 타당도, 반응도, 해석가능성을 입증하여야 한다. 이를 통해 저자는 자료 추출 및 합성에 사용할 도구를 결정할 수 있다. 예를 들어 다음 도구들은 통증 치료를 위한 중재의 유용성을 평가하기 위해 연구자가 일차 연구에서 사용할 수 있는 검증된 환자보고 통증 도구이다.

- 7-item Integrated Pain Score
- 10-point Visual Analogue Scale for Pain
- 20-item McGill Pain Questionnaire
- 56-item Brief Pain Inventory

일부 임상 분야에서는 핵심 중재결과 세트(core outcome set)를 통해 적절한 PRO가 사용되었는지를 판단하게 된다. 류마티스학 분야에서 Outcome Measures in Rheumatology(OMERACT) 이니셔티브는 핵심 영역 세트와 핵심 중재결과 세트를 확인하기 위해 OMERACT Filter 2.0으로 알려진 개념 틀을 개발하였다. 이것은 일반적인 틀이기 때문에 류마티스 분야 외에 다른 분야에서도 핵심 중재결과 세트 개발에 적용할 수 있다.

사전에 정의된 계층구조의 예로, 무릎 골관절염에 대해 OMERACT는 체계적 문헌고찰을 수행하기 위해 통증 및 신체기능을 평가하는 PROM를 반응도를 기반으로 계층구조 만들어 사용하였다(Juhl et al 2012).

저자는 계층 구조에 포함되지 않는 PROM을 제외할지 또는 선호하는 방법을 사용할 수 없는 경우 추가 조치를 포함할지 여부를 미리 결정해야 한다.

27.4.2 다양한 PROM으로부터 자료 합성

계층구조는 문헌고찰 저자가 선호하는 측정도구를 확인하고 부적절한 것으로 간주되는 일부 측정도구를 제외하는데 도움이 될 수 있지만, 저자는 한 연구 내에서 또는 여러 연구에서 주어진 구성을 측정하기 위해 여러 가지 PROM을 사용하는 연구에 직면할 가능성이 높다. 이때 다양한 측정도구의 합성에 접근하는 방법을 결정해야 하며, 그 중에서도 특정 구성에 대한 단일 메타분석에 포함할 측정도구를 고려해야 한다(Christensen et al 2015).

통계적 합성이 적절한지 결정할 때, 문헌고찰 저자는 사용된 PROM의 기초 구성에 대한 정확한 개념을 얻기 위해 행간을 읽는 경우가 종종 있다. 때때로 PROM의 개발 논문이나 이전 사용 논문을 참조하거나 측정되는 개념을 이해할 수 있는 다른 도구를 살펴야 할 수도 있다.

동일한 메타 분석에서 서로 다른 PROM을 포함하는 것이 적절하다는 것과 관련하여 가장

강력한 근거는 도구들 간의 실질적인 상관성을 발견하는 것이다. 예를 들어 COPD 환자에서 HRQoL을 측정하는데 사용되는 두 가지 주요 도구는 Chronic Respiratory Questionnaire(CRQ)와 St. George's Respiratory Questionnaire(SGRQ)이다. 개별 연구에서 두 가지 설문지 사이의 상관성은 횡단(한 지점에서의 상관관계)과 종단(변화의 상관관계) 비교에서 모두 0.3에서 0.6까지 다양했다.

어떤 PRO와 PROM이 메타 분석에 포함될 지 결정한 후, 문헌고찰 저자는 자신이 제시한 연구결과가 목표 독자에게 해석될 수 있도록 보장해야하는 어려움에 직면하게 된다. 예를 들어 CRQ를 사용한 일련의 무작위배정 임상시험에서 재활과 표준치료의 평균 차이가 1.0(95% CI 0.6~1.5) 이라고 말하면 많은 독자는 이것이 사소하고 작지만 중요한지, 중등도인지 큰 효과인지를 확신하지 못할 것이다. 비슷하게, 표준화된 평균 차이(standardized mean difference)에 대한 해석은 대부분 도전적이다.

27.5 참고문헌과 읽을거리

- Boers M, Kirwan JR, Wells G, Beaton D, Gossec L, d'Agostino MA, Conaghan PG, Bingham CO, 3rd, Brooks P, Landewe R, March L, Simon LS, Singh JA, Strand V, Tugwell P. Developing core outcome measurement sets for clinical trials: OMERACT filter 2.0. Journal of Clinical Epidemiology 2014; 67: 745-753.
- Christensen R, Maxwell LJ, Jüni P, Tovey D, Williamson PR, Boers M, Goel N, Buchbinder R, March L, Terwee CB, Singh JA, Tugwell P. Consensus on the Need for a Hierarchical List of Patient-reported Pain Outcomes for Metaanalyses of Knee Osteoarthritis Trials: An OMERACT Objective. Journal of Rheumatology 2015; 42: 1971-1975.
- COMET. Core Outcome Measures in Effectiveness Trials 2018. http://www.cometinitiative.org.
- FDA. Guidance for Industry: Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. Rockville, MD; 2009. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf.
- FDA. Clinical Outcome Assessment Program Silver Spring, MD: US Food and Drug Administration; 2018. https://www.fda.gov/Drugs/DevelopmentApprovalProcess/DrugDevelopmentToolsQualificationProgram/ucm284077.htm.
- Guyatt GH, Naylor CD, Juniper E, Heyland DK, Jaeschke R, Cook DJ. Users' guides to the medical literature. XII. How to use articles about health-related quality of life. EvidenceBased Medicine Working Group. JAMA 1997; 277: 1232-1237.
- Johnston BC, Patrick DL, Devji T, Maxwell LJ, Bingham III CO, Beaton D, Boers M, Bri-

el M, Busse JW, Carrasco-Labra A, Christensen R, da Costa BR, El Dib R, Lyddiatt A, Ostelo RW, Shea B, Singh J, Terwee CB, Williamson PR, Gagnier JJ, Tugwell P, Guyatt GH. Chapter 18: Patient-reported outcomes. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.0 (updated July 2019). Cochrane, 2019. Available from www.training.cochrane.org/handbook.

- Juhl C, Lund H, Roos EM, Zhang W, Christensen R. A hierarchy of patient-reported outcomes for meta-analysis of knee osteoarthritis trials: empirical evidence from a survey of high impact journals. Arthritis 2012; 2012: 136245.

부록

관련 용어 번역



A

accuracy	정확도
assumed comparator risk	추정비교군위험도
active treatment	활성치료
adherence	준수
adhering	준수
adjusted	보정(된)
adjusted effect estimates	보정 효과 추정치
administrative units	행정단위
adverse	유해
adverse drug reaction	유해 약물반응
adverse effects	유해효과
adverse events	유해사건
advisory group	자문단
aggregation	집계
all-cause mortality	모든 원인 사망률
allocation	배정
answerable	대답가능한
applicability	적용성
arm level	군 수준
arm-based	군 기반
assessment	평가
assignment	배정
assumed	추정(된)
as-treated	치료받은 대로
at-intervention	중재중
attrition	탈락
auditability	감사가능성
author team	저자 팀
away from null	무효점에서 멀어짐

B

baseline	기저, 기저치, 기저상태
baseline observation carried forward, bOCF	기저상태 관찰 이월
baseline risks	기저 위험도
bayesian synthesis	베이지안 합성
benefit	이득, 편익 (경제성 평가시)
bespoke	맞춤형
bias	비뚤림
bias arising from the randomization process	무작위과정에서 생기는 비뚤림
bias due to deviations from intended interventions	의도한 중재에서 일탈로 인한 비뚤림
bias due to missing outcome data	중재결과 자료의 결측으로 인한 비뚤림
bias in measurement of the outcome	중재결과 측정의 비뚤림
bias in selection of the reported result	보고된 연구결과 선택의 비뚤림
biased	편향된
bibliographic database indexing systems	서지 데이터베이스 색인 시스템
biomarker	생표지자
biomedical electronic literature database	전자 생의학 데이터베이스
blinding	눈가림
body of evidence	전체적 근거
boolean	불리언
box-and-whisker plots	상자수염그림
bubble plots	버블플롯

C

capture-recapture	포획-재포획
case-control study	환자-대조군 연구
catalogue	목록
causal chain	인과 사슬
causal chain analysis	인과 사슬 분석
censored data	중도절단 자료
certainty of the evidence	근거수준
clinical study report	임상연구보고서
cluster	군집
cluster-randomized trial	군집무작위임상시험
cochrane review(s)	코크란리뷰
coding	코딩
coherence	일관성
co-intervention	공동 중재
comment	논평
common-effect	공통효과
comparable	유사한
comparator	비교중재
comparator intervention	비교 중재
comparison	비교
complete case	완료 증례
complete-case analyses	완료증례 분석
complex intervention	복합중재
component based	구성요소 기반
composite outcomes	복합 중재결과
comprehensiveness	포괄성
conceptual framework	개념적 틀
condition	상태
conduct bias	수행 비뚤림
conference abstract	학술대회 초록
confidence limits	신뢰 한계
confirmatory	확증적
conflicts of interest	이해상충
confounder	교란요인
confounding	교란
construct	구성
consumer	소비자
context	맥락
contextual	맥락적
contextual variations	맥락적 변이
contrast-based'	대비 기반
control	통제
control group, arm	대조군
controlled vocabulary	통제어
convergent mixed-methods review	수렴 혼합방법 문헌고찰
core outcome sets	핵심 중재결과 세트
correspondence	교신
corresponding intervention risk	해당 중재의 위험도
cost benefit analysis	비용편익분석
cost effective analysis	비용효과분석
cost of illness	질병비용
cost utility analysis	비용효용분석
count or rate data	계수자료, 비율 자료
covariates	공변량
credibility	신뢰성
criteria for study eligibility	연구 적격 기준
critical	매우 높음(rob평가), 핵심적(형용사)
critical outcome	핵심적 중재결과
cross-over	교차
cross-over trial	교차 임상시험
currency	현재성

D, E

data	자료
data extractor	자료 추출자
database	데이터베이스
day care	보육
design features	설계 특성
detection bias	결과 확인 비뚤림
deterministic assignment	결정론적 배정
deviation(s)	이탈, 벗어나
diagnostic detection bias	진단 결과 확인 비뚤림
diagnostic test bias	진단 검사 비뚤림
diagram	다이어그램
dichotomous	이분형
differential	차등적
differential measurement error	차등적 측정 오류
differential misclassification	차등적 오분류
disadvantage	취약 계층
distiller SR	distillerSR
divergence	다양성
diversity	다양성
domain	영역
domain-based tools	영역기반도구
dose-response gradient	용량 - 반응 변화
double blinding	이중 눈가림
downgrading	하향평가
drop out	탈락
economics	경제성
economics components	경제성 구성 요소
effect	효과
effect modification	효과변경
effect modifier	효과 변경요인
effect of adherence to intervention	중재준수효과
effect of assignment	배정 효과
effect of interest	관심 효과
eligibility	적격성
eligibility criteria	적격기준
eligible	적격
eligible studies	적격 연구
endpoint	종결점
endpoint outcome	최종 중재결과
end-user	최종 사용자
enrolment personnel	모집 담당자
epistemology	인식론
equal-effect	동등효과
equity	형평성
errata	정정기사
error	오류
evaluation	평가
event-free survival	무사건 생존
exclusion	배제
experimental context	임상시험의 맥락
expermental interventions	실험 중재
expertise	전문성
exploratory	탐색적
extraction of data	자료추출

F, G, H

factorial	팩토리얼
favour	선호
Favours experimental	실험 중재 선호

feasible	실현가능한
finding	연구결과
fitness for purpose	목적 적합성
fixed-effect	고정 효과
flow diagram	흐름도
focus	초점
focus group	포커스 그룹
follow up	추적
forest plot(s)	숲그림
formatting	구성하기
forward citation tracking	전방 인용 추적
frame	틀
framework	틀
framing data	자료 형태화
fraudulent studies	부정한 연구
framework	틀
freetextbox	프리텍스트 박스
free-text	프리텍스트
full economic evaluations	완전 경제성 평가
funder	자금제공자
funnel plot	깔대기 그림
gap	간극
generic instrument	일반적 도구
grant database summary	연구비 지원 데이터베이스 요약
grey literature	회색 문헌
handbook	핸드북
harm	위해
harvest plots	하베스트 플롯
hazard ratio	위험비
health decision maker	보건의료 의사결정자
health economist	보건경제학자
health-related quality of life	건강 관련 삶의 질
theoretical sampling	이론 표본추출
heterogeneity	이질성
highly desirable	권고
hybrid	하이브리드

I, J, K

illustrative risk	예시적 위험
immortal time bias	불멸의 시간 비뚤림
implementation	실행
implications	영향
imputation methods	결측값 대치 방법
impute	결측값대치
inception bias	개시 비뚤림
inception cohorts	개시 코호트
inclusion	포함
Inconsistencey	비일관성
incremental	증분
incremental beneficial effect	증분 이득 효과
index terms	색인어
indexing	색인
indication	적응증
indirect comparison	간접비교
individual participant data	개별 연구대상자 자료
Individually randomized trial	개별 무작위 임상시험
inequity	불평등
information specialist	정보 전문가
informative censoring	정보중도절단
in-house databases	사내데이터베이스
instrumental variable analyses	도구 변수 분석

integrated full systematic review	통합적 완전 체계적 문헌고찰
Integrity	진실성
intended effects	의도한 효과
intended interventions	의도한중재
intended to treat	배정된대로
intensity sampling	강도 표본추출
intention to treat analysis	배정된대로분석
intention-to-treat' effect	중재배정 효과
interaction	상호작용
interval-censored survival	간격 중도절단 생존
intervention	중재
intervention complexity	중재 복합성
intervention effect	중재효과
inverse probability weighting	역 확률 가중치
inverse variance	역분산
investigators	연구자
journal article	학술지 논문
key concept	핵심 개념
key foundings	핵심사항
key points	핵심포인트
known unknowns'	알려진 무지

L, M

label	라벨
language bias	언어비뚤림
last observation carried forward, LOCF	최종 관찰 이월
lead author	주저자
lead time bias	시간 지연 비뚤림
legal examination	법률적 검토
librarian	사서
likelihood-based	우도 기반
literature surveillance services	문헌 감시 서비스
local databases	지역 데이터베이스
logic model	로직 모델
lost to follow up	추적관찰 소실
low	낮음
lumping	광범위
machine learning	기계학습
magnitude of effect	효과크기
major adverse cardiac and cerebrovascular events(mACCE)	주요 심장 및 뇌혈관 유해사건
major morbidity	주요 이환(상태)
mandatory	필수
mantel Haenszel	맨텔-핸젤
mapping	매핑
marginal structural models	한계 구조 모델
masking	눈가림
maximum variation sampling	최대 변동 표본추출
measurement error	측정 오류
measures	측정치
mechanism of action	작용기전
mediator	중재자
meta-analytic path analysis	메타 분석적 경로 분석
meta-ethnography	메타 민속지학
meta-regression	메타 회귀 분석
metric	메트릭
minimization	최소화
misclassification	오분류
missing	결측
missing data	결측치

missing not at random, mNAR	무작위적으로 결측되지 않음
missing outcome data	중재결과 자료 결측
missing results	연구결과 결측
missingness	결측성, 결측여부
mixed-methods synthesis	혼합방법합성
model-based meta-analysis	모델 기반 메타 분석
moderate	중등도
moderator	조정자
modified ITT	수정된 ITT
moment-based	모멘트 기반
multicomponent interventions	다요소 중재
multicomponent reviews	다요소 문헌고찰
multiple treatment attempts	다중처리시도

N, O

naïve analysis	순수 분석
narrative	서사
narrative data	서사자료
natural units of measurement	자연 측정 단위
network diagrams	네트워크다이어그램
network meta-analysis	네트워크 메타분석
no	아니다.
non-differential	비차등적
non-differential misclassification	비차등적 오분류
non-informative	비정보적
non-protocol interventions	비프로토콜 중재
non-public sources	비공식적 자료원
non-randomized	비무작위
non-randomizedstudiesofinterventions(NRSI)	중재비무작위연구
non-reporting bias	비보고 비뚤림
null	무효점, 귀무
number needed to treat (NNT)	필요 치료 수
odds ratio	오즈비
online technical supplement	온라인 부록
outcome(s)	중재결과
outlines	개요
overall risk of bias	전반적 비뚤림 위험
over-ascertainment	과다확인
overview of reviews	문헌고찰 overview

P, Q

Parallel	평행
parallel-group trial	평행군 임상시험
partial economic evaluations	부분 경제성 평가
participant(s)	연구대상자
participant-reported outcomes	연구대상자 보고 중재결과
peer review	동료검토
per protocol effect	프로토콜 준수 효과
per protocol, PP	프로토콜 대로
period effect	주기효과
per-protocol analysis	프로토콜 대로 분석
plausible confounding	그럴듯한 교란
point estimates	점 추정
policy maker	정책 결정자
population	인구집단
post-baseline deviations	기저 이후 이탈

post-randomization exclusions	무작위배정 후 배제
pragmatic randomized trial	실용 무작위배정 임상시험
precision	정밀도
preconception	선입견
prediction interval	예측 구간
pre-intervention	중재전
prespecification	사전 정의
prevalent user	기존 사용자
primary beneficial outcome	일차 이득 중재결과
priori theory	선험적 이론
process	과정
process-oriented	과정 지향
prognostic factors	예후 요인
prognostic study	예후 연구
promising	유망한
protocol	프로토콜
proxy measure	대리 측정치
publication bias	출판비뚤림
purposive sampling	목적적 표본추출
qualitative comparative analysis	질적 비교 분석
qualitative evidence synthesis	질적 근거합성
qualitative research	질적 연구
quasi-experimental	준실험적

R

random permuted blocks	무작위 치환 블록
random-effect	변량 효과
randomization process	무작위배정 과정
randomized trial	무작위배정 임상시험
rarely randomized studies	준무작위 연구
rate ratio	발생비
rationale	타당성
recall bias	회상 비뚤림
references	참고문헌
reflection	반영
register	등록부
registry	레지스트리
registry record	레지스트리 자료
regression coefficients	회귀 계수
regulatory documents	규제문서
regulatory reviews	규제 문헌고찰
relative recall technique	상대적 회상법
reliability	신뢰도
replicable	재현가능한
replication	복제
reporting bias	보고 비뚤림
research ethics application	IRB 지원 서류
research report	연구 보고서
resources	자원
responsiveness	반응도
result(s)	연구 결과
retracted publications	취소 출판물
retractions	논문 취소
retrospective	후향적
review	문헌고찰
review author	문헌고찰 저자
review question	문헌고찰 질문
review team	문헌고찰 팀
reviews	문헌고찰
rigour	엄격성
risk difference	위험차

risk of bias	비뚤림 위험
risk ratio	위험비
risk to rigour tool	엄격성 위험 평가도구
robustness	견고성

S

sample	표본
scoping review	주제범위 문헌고찰
scoping searches	주제범위 검색
selection bias	선택 비뚤림
selective dissemination	선택적 확산
self-completed questionnaires	자기기입 설문지
semi parametric	반모수
sensitivity	민감도
sequence generation	순서생성
sequential reviews	순차적 문헌고찰
serious	높은
serious adverse event	중대한 유해 반응
setting	세팅
side effect	부작용
signalling question	신호 질문
some concerns	일부 우려
sources	출처
sources to search	검색원
split	좁은
stakeholder	이해당사자
standard care	표준 치료
standardized subject term	표준화 주제어
stepped-wedge trials	스텝웨지임상시험
stratified randomization	층화 무작위 배정
structural equation modelling	구조 방정식 모델링
structuring	구조화
study design	연구 설계
study filter	연구 필터
study-based	연구기반
subgroup	하위군
subgroup analysis	하위군 분석
subject heading, MesH	MesH
subscales	하위척도
subset	하위세트
summary of findings	결과요약표
supplementarycriteria	보완기준
surrogate	대리결과
surrogate markers	대리 표지자
synergistic	상승
synthesis	합성(과정, 결과)
systematic error	체계적 오류
system-based	시스템 기반

T, U

target trial	목표 임상시험
tele-home care	원격 재택 치료
test	검정
text mining approaches	텍스트 마이닝 접근법
text word	텍스트워드
thematic synthesis	이론 합성
theme	주제
theories of change	변화 이론
thick	심층적
thin	빈약한

threshold	임계치
time-to-event	사건발생 시간
time-to-event analyses	사건발생 시간 분석
time-to-event data	사건발생 시간 자료
towards null	무효점을 향함
transferability	전이성
transitivity	이행성
transparency	투명성
treatment effect	치료효과
trial investigators	임상시험연구자
trial participants	연구대상자
trial registers	임상시험 등록처
trial sibling	형제 임상시험
trial registrations	임상시험 등록
trialist	임상시험 연구자
trichotomous	삼분형
true value	참값
type of data	자료유형
unadjusted intervention effect estimate	비보정 중재 효과 추정치
under-ascertainment	과소확인
underlying value	기초값
unit of interest	관심 단위
unknown unknowns'	알려지지 않은 무지
update	갱신
usual practice	통상진료
utility	효용

V, W, Y, Z

validity	타당도
variation	변이
vote counting	다수결 집계
workshop	워크숍
yes/probably yes	그렇다/아마도 그렇다
zero events	사건발생 없음